Omri Sarig

Horocycle flows on surfaces with infinite genus

Geometric and Ergodic Aspects of Group Actions

TIFR, April 2015

Contents

Preface

We discuss the ergodic theory of horocycle flows on hyperbolic surfaces with infinite genus. In this case all finite invariant measures are trivial (they are all carried by closed orbits encircling cusps), and the interesting ergodic phenomena happens on the level of infinite invariant Radon measures.

Chapter 1 gives a recipe for constructing such measures. If the surface is big enough to accommodate a non-constant positive eigenfunction for the Laplacian, then this eigenfunction can be used to write down an explicit formula for a non-trivial horocycle invariant Radon measure ("Babillot's construction"). The main result of the chapter is that for a large class of surfaces *all* invariant Radon measures arise this way, and extremal eigenfunctions lead to ergodic measures.

Chapter 2 discusses some ergodic theoretic features of these measures. A generalized law of large numbers (GLLN) for an ergodic infinite invariant measure is a procedure which accepts as input a record of the times an orbit spends inside a set E, and gives as output the measure of E. An important pathology in infinite ergodic theory is that some ergodic invariant measures do not possess GLLN. This is demonstrated in the particular case of horocycle flows on $\mathbb{Z}^d$-covers of compact surfaces: Although such surfaces possess infinitely many different ergodic invariant Radon measures, only one, the volume measure, has a GLLN.

Chapter 3 is concerned with the problem of equidistribution. An important feature of infinite genus is the coexistence of many different globally supported ergodic infinite invariant measures. Each has its own set of generic points (points of validity for the ratio ergodic theorem for all continuous functions with compact support). The main result is the description of the generic points for the different ergodic invariant measures for horocycle flows on a $\mathbb{Z}^d$-cover of a compact surface.

Chapter 4 gives a sketch of the proof of the result mentioned in chapter 1, that every ergodic invariant Radon measure for the horocycle flow on a tame surface arises from an extremal positive eigenfunction via Babillot's construction. For reasons of exposition, we only give the proof in the very tame case.

Acknowledgements. This set of notes constituted the basis for a series of lectures given in April 2015 as part of the program "Geometric and ergodic aspects of group actions", at the Tata Institute for Fundamental Research, Mumabi. The author would like to thank the organizers of the program and the staff of TIFR for the kind hospitality. The author acknowledges the support of ISF grant 1149/18.

Chapter 1
Classification of invariant Radon measures

This chapter discusses the relation between infinite, locally finite, invariant measures for the horocycle flow of a hyperbolic surface M of infinite genus, and positive eigenfunctions of the Laplacian on M.

1.1 Basic definitions

Hyperbolic surfaces: The *Hyperbolic plane* has two classical models: The first is the *Poincaré disc* $\mathbb{D} := \{z \in \mathbb{C} : |z| < 1\}$, together with the Riemannian metric $2|dz|/(1-|z|^2)$. The second is the *upper half plane* with the Riemannian metric $|dz|/\mathrm{Im}(z)$. The most basic fact on these models is that their sets of orientation preserving isometries are, respectively, $\text{Möb}(\mathbb{D}) := \{\text{Möbius maps which preserve } \mathbb{D}\}$ and $\text{Möb}(\mathbb{H}) := \{\text{Möbius maps which preserve } \mathbb{H}\}$.

A *hyperbolic surface* is a Riemannian surface M s.t. every $p \in M$ has a neighborhood isometric to some open subset of the hyperbolic plane.

A hyperbolic surface is called *complete*, if every geodesic ray can be extended indefinitely. *Throughout these notes, unless stated otherwise, all surfaces are complete, connected, and orientable.* A classical result (the Killing-Hopf Theorem) says that in this case there is a discrete subgroup $\Gamma \subset \text{Möb}(\mathbb{D})$ without elements of finite order s.t. M is isometric to the *orbit space* of Γ

$$\Gamma \setminus \mathbb{D} := \{\Gamma z : z \in \mathbb{D}\}\ ,\ \Gamma z := \{\varphi(z)\}_{\varphi \in \Gamma}$$

together with the metric induced by the covering map $p : \mathbb{D} \to \Gamma \setminus \mathbb{D}$, $p(z) = \Gamma z$. For a nice account of the Killing-Hopf theorem, see [Sti].

The geodesic flow: The *geodesic flow* is the flow $g : T^1M \to T^1M$ on the unit tangent bundle $T^1M := \{\mathbf{v} \in TM : \|\mathbf{v}\| = 1\}$ which moves $\mathbf{v} \in T^1M$ at unit speed along its geodesic. Completeness guarantees that $g^t(\mathbf{v})$ exists for all t.

If $M = \mathbb{D}$, then g^t moves $\mathbf{v}$ along the unique arc of a circle or line which is tangent to $\mathbf{v}$ and perpendicular to $\partial\mathbb{D}$. This is obvious for the vector $\mathbf{v}_0$ based at the origin of $\mathbb{D}$ which points to the right. For other vectors $\mathbf{v}$, find a suitable isometry $\varphi \in \text{Möb}(\mathbb{D})$ s.t. $\mathbf{v} = \varphi(\mathbf{v}_0)$, and recall that Möbius maps preserve angles and map lines or circles to line or circles.

If $M = \Gamma \setminus \mathbb{D}$, we can calculate $g^t(\mathbf{v})$ for $\mathbf{v} \in T^1M$ by first lifting $\mathbf{v}$ to $T^1\mathbb{D}$, applying the geodesic flow of $\mathbb{D}$ to the lift, and projecting the result to T^1M.

Horocycle flow: The *stable horocycle* of a unit tangent vector is the strong stable manifold of $\mathbf{v}$ with respect to the geodesic flow:

$$\text{Hor}(\mathbf{v}) := W^{ss}(\mathbf{v}) := \{\mathbf{u} \in T^1M : \text{dist}\,(g^s(\mathbf{u}), g^s(\mathbf{v})) \xrightarrow[s\to\infty]{} 0\}.$$

It is a fact that $\text{Hor}(\mathbf{v})$ is a smooth one-dimensional curve in T^1M. The *horocycle flow* $h : T^1M \to T^1M$ is the flow which moves $\mathbf{v} \in T^1M$ at unit speed and in the positive direction its stable horocycle.

The simplest possible case is when the surface is $\mathbb{D}$. In this case $\text{Hor}(\mathbf{v})$ is made of all inward pointing vectors orthogonal to a circle which touches $\partial\mathbb{D}$ at one point (figure 1.1). This is easiest to see in the upper half plane model, first for vectors $\mathbf{v}$ pointing "up", and then for general vectors using an isometry.

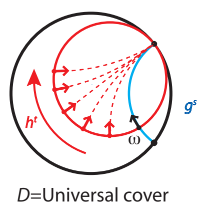

Fig. 1.1 The horocycle flow on $T^1\mathbb{D}$

The horocycle flow for general surface $\Gamma \setminus \mathbb{D}$ can be calculated by lifting to $\mathbb{D}$, applying the flow there, and projecting the result.

Ergodic invariant Radon measures (e.i.r.m.): A Borel measure m on T^1M is

- *ergodic* (for h) if every h–invariant function is equal to a constant function m–a.e.,
- *invariant* (for h) if $m \circ h^t = m$ for all $t \in \mathbb{R}$,
- *Radon* if every compact set has finite measure (non-compact sets are allowed to have infinite measure). Equivalently, m is Radon if every continuous function with compact support is integrable.

When M has finite genus, the horocycle flow has, up to normalization, exactly one e.i.r.m. which is not carried by a single orbit. But this phenomenon breaks down in infinite genus, as can be seen in the following table.

	h–ergodic invariant Radon measures	**extremal positive eigenfunctions of Δ**
compact (Furstenberg)	volume measure	constant
finite area (Dani, Smillie)	volume measure + cusp periodic orbits	constant + Eisenstein series
finite genus (Burger, Roblin)	Burger–Roblin measure + cusp periodic orbits + orbits escaping to funnels	Patterson's function + Eisenstein series + Eisenstein-Patterson series
$\mathbb{Z}^d$–covers (Babillot & Ledrappier)	d–parameter family "BL measures"	d–parameter family Lin & Pinchover
polycyclic* covers of exponential growth (Ledrappier & Sarig)	there exists an e.i.r.m $\neq$volume which is invariant under the horocycle flow *and* the geodesic flow	there exists a positive unbounded harmonic function (Bougerol & Élie)

* A group G is *polycyclic* if $G = G_n \triangleright \cdots \triangleright G_0 = \{1\}$ and G_i/G_{i-1} is cyclic.

Notice the similarity between the list of e.i.r.m. and the list of extremal positive eigenfunctions of the Laplacian. Babillot, who was the first to observe this, suggested a mechanism for producing invariant measures out of positive eigenfunctions, and conjectured that all measures arise this way.

1.2 Invariant measures arising from positive eigenfunctions

1.2.1 kan coordinates

Given $e^{i\theta} \in \partial\mathbb{D}$, let $\omega(e^{i\theta}) \in T^1\mathbb{D}$ be the unit tangent vector based at the origin and pointing at $e^{i\theta}$. Every unit tangent vector $\mathbf{v} \in T^1\mathbb{D}$ can be written uniquely as

$$\mathbf{v} = (h^t \circ g^s)[\omega(e^{i\theta})].$$

The *kan–coordinates* of $\mathbf{v}$ are $e^{i\theta}, s$ and t (*kan* is pronounced *kay-ay-en*).

Here is the description of the horocycle, geodesic and Möbius actions in *kan* coordinates:

Theorem 1.1. *In the kan-coordinate system*

1. $h^t(e^{i\theta_0}, s_0, t_0) = (e^{i\theta_0}, s_0, t_0 + t)$
2. $g^s(e^{i\theta_0}, s_0, t_0) = (e^{i\theta_0}, s_0 + s, t_0 e^{-s})$.
3. For every $\varphi \in \text{Möb}(\mathbb{D})$,

$$\varphi(e^{i\theta_0}, s_0, t_0) = (\varphi(e^{i\theta_0}), s_0 - \log|\varphi'(e^{i\theta_0})|, t_0 + \textit{something independent of } t_0) \qquad (*)$$

Part 1 is obvious. Part 2 is because $g^s \circ h^{t_0} = h^{t_0 e^{-s}} \circ g^s$. Part 3 is proved in the appendix.

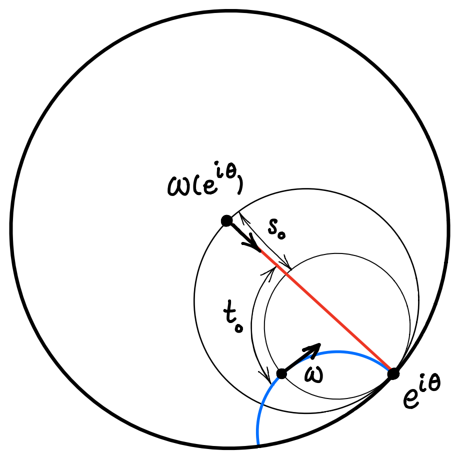

Fig. 1.2 The *kan*–coordinates $(e^{i\theta_0}, s_0, t_0)$ of $\omega \in T^1\mathbb{D}$

1.2.2 The Laplacian and its eigenfunctions

The Laplacian: The Laplacian on $\mathbb{D}$ is a second order differential operator $\Delta_{\mathbb{D}}$ on $\mathbb{D}$ which commutes with all hyperbolic isometries: $(\Delta_{\mathbb{D}}F)\circ\varphi = \Delta_{\mathbb{D}}(F\circ\varphi)$ for all $\varphi \in \mathrm{Isom}(\mathbb{D})$. This determines $\Delta_{\mathbb{D}}$ up to a scalar.

The standard choice, in coordinates, is $\Delta_{\mathbb{D}} = \frac{1}{4}(1-x^2-y^2)^2\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)$ for the unit disc, and $\Delta_{\mathbb{H}} = y^2\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)$ for the upper half-plane.

Every function $F(\Gamma z)$ on $\Gamma\setminus\mathbb{D}$ can be identified with the Γ-invariant function on $\mathbb{D}$ given by $\widetilde{F}(z) := F(\Gamma z)$. Since $\Delta_{\mathbb{D}}$ commutes with Möb($\mathbb{D}$), the Laplacian of a Γ–invariant function on $\mathbb{D}$ is Γ–invariant, therefore $\Delta_{\mathbb{D}}$ descends to a well–defined second order differential operator $\Delta_{\Gamma\setminus\mathbb{D}}$ on $\Gamma\setminus\mathbb{D}$, $(\Delta_{\Gamma\setminus\mathbb{D}}F)(\Gamma z) := (\Delta_{\mathbb{D}}\widetilde{F})(z)$. We will often abuse notation and write Δ for $\Delta_{\mathbb{D}}$ or $\Delta_{\Gamma\setminus\mathbb{D}}$.

Positive eigenfunctions: A C^2–function $F : \Gamma\setminus\mathbb{D}\to\mathbb{R}$ will be called an *eigenfunction*, if $\Delta F = \lambda F$. *We allow infinite L^2–norm.* We will be interested in positive eigenfunctions.

The most important eigenfunction on $\mathbb{D}$ is *Poisson's kernel*:

$$P(e^{i\theta}, z) = \frac{1-|z|^2}{|e^{i\theta}-z|^2} \qquad (e^{i\theta}\in\partial\mathbb{D}, |z|<1)$$

Theorem 1.2. *$P(e^{i\theta},z)^\alpha$ is a smooth, positive, unbounded function on $\mathbb{D}$. The level sets $P(e^{i\theta},\cdot)^\alpha = \text{const.}$ are horocycles which touch $\partial\mathbb{D}$ at $e^{i\theta}$. $\Delta P(e^{i\theta},\cdot) = 0$; and $\Delta P(e^{i\theta},\cdot)^\alpha = \lambda P(e^{i\theta},\cdot)$, where $\lambda = \alpha(\alpha-1)$.*

Proof. This is best seen in the upper half-plane model. Here $\Delta_{\mathbb{H}} = y^2(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2})$, so $f(x+iy) = y^\alpha$ is a positive eigenfunction with eigenvalue $\lambda = \alpha(\alpha-1)$. The level sets $y^\alpha = \text{const.}$ are horocycles tangent to $\mathbb{H}$ at ∞.

To pass to $\mathbb{D}$ we use the isometry $\varphi : \mathbb{D} \to \mathbb{H}$, $\varphi(z) := i\frac{1+z}{1-z}$ and deduce that $f(\varphi(z)) = [\mathrm{Re}(\frac{1+z}{1-z})]^\alpha = P(1,z)^\alpha$ is a positive eigefunction on $\mathbb{D}$ with eigenvalue λ. $\Delta_{\mathbb{D}}$ commutes with Möb($\mathbb{D}$), and in particular with $\psi(z) = e^{-i\theta}z$. So $P(e^{i\theta},z) = P(1,\psi(z))$ a positive eigefunction on $\mathbb{D}$ with eigenvalue λ. □

A positive eigenfunction is called *extremal* if it is not a non-trivial average of positive eigenfunctions with the same eigenvalue which are not proportional to it. Extremal eigenfunctions are sometimes called *minimal*, because the extremality of F is equivalent to the following property: any positive eigenfunction G with the same eigenvalue as F such that $0 \le G \le F$ is proportional to F.

It turns out that $P(e^{i\theta},z)^\alpha$ is extremal whenever $\alpha \ge \frac{1}{2}$, and these eigenfunctions suffice to represent all other positive eigenfunctions:

Theorem 1.3 (Karpelevich Representation Theorem). *Every positive eigenfunction on $\mathbb{D}$ has eigenvalue $\lambda \ge -\frac{1}{4}$, and admits a unique representation of the form*

$$F(\cdot) = \int_{\partial\mathbb{D}} P(e^{i\theta},\cdot)^\alpha d\nu(e^{i\theta}) \tag{1.1}$$

where ν is a finite positive measure on $\partial\mathbb{D}$, $\alpha \ge \frac{1}{2}$, and $\alpha(\alpha-1) = \lambda$.

Remark: Sometimes there are other representations with $\alpha < \frac{1}{2}$, see [Ba].

This theorem treats positive eigenfunctions on $\mathbb{D}$. If we want eigenfunctions on $\Gamma\setminus\mathbb{D}$, then we need to choose ν to make the right hand side of (1.1) is Γ–invariant. A finite positive measure ν on $\partial\mathbb{D}$ is called *Γ–conformal with parameter α* if

$$\frac{d\nu\circ\varphi}{d\nu} = |\varphi'|^\alpha \quad \text{for every } \varphi \in \Gamma. \tag{1.2}$$

These measures were introduced by Patterson and Sullivan in the seventies. Sullivan called these measures "conformal densities."

Theorem 1.4 (Sullivan). *Fix $\alpha \ge \frac{1}{2}$, and suppose $F(z)$ and ν are related by (1.1). $F(z)$ is Γ–invariant iff ν is Γ-conformal with parameter α.*

Proof. The proof uses the *harmonic measures* $d\lambda_z := P(e^{i\theta},z)d\lambda$ where λ is Lebesgue's measure on $\partial\mathbb{D}$.

Since $\Delta_{\mathbb{D}} = \frac{1}{4}(1-|z|^2)^2(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2})$, harmonic functions for $\Delta_{\mathbb{D}}$ are the same as harmonic functions for $\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$. Every $g \in C(\partial\mathbb{D})$ determines a unique harmonic function $G(z)$ with boundary values $G|_{\partial\mathbb{D}} = g$. G is given by the Poisson Integral

Formula $G(z) = \int_{\partial\mathbb{D}} g d\lambda_z$. For every $\varphi \in \text{Möb}(\mathbb{D})$, $G(\varphi(z))$ is harmonic, because if G is the real part of a holomorphic function, so is $G \circ \varphi$. Since $G \circ \varphi|_{\partial\mathbb{D}} = g \circ \varphi$, $\int_{\partial\mathbb{D}} g \circ \varphi d\lambda_z = \int_{\partial\mathbb{D}} g d\lambda_{\varphi(z)}$. The last identity holds for all $g \in C(\partial\mathbb{D})$. Therefore, we have the *harmonic measures identity* $\lambda_{\varphi(z)} = \lambda_z \circ \varphi^{-1} \quad (\varphi \in \text{Möb}(\mathbb{D}))$. This translates to the following identity for Poisson kernels:

$$P(e^{i\theta}, \varphi(z)) = P(\varphi^{-1}(e^{i\theta}), z) \cdot |(\varphi^{-1})'(e^{i\theta})|. \tag{1.3}$$

We can now prove the theorem. Suppose $F(z) = \int_{\partial\mathbb{D}} P(e^{i\theta}, z)^\alpha d\nu(e^{i\theta})$, then $F(\varphi(z)) = \int_{\partial\mathbb{D}} P(e^{i\theta}, \varphi(z))^\alpha d\nu(e^{i\theta}) = \int_{\partial\mathbb{D}} P(\varphi^{-1}(e^{i\theta}), z)^\alpha \cdot |(\varphi^{-1})'(e^{i\theta})|^\alpha d\nu(e^{i\theta}) = \int_{\partial\mathbb{D}} P(e^{i\eta}, z)^\alpha |(\varphi^{-1})'(\varphi(e^{i\eta}))|^\alpha \frac{d\nu\circ\varphi}{d\nu} d\nu(e^{i\eta})$. Thus $F(\varphi(z)) = F(z)$ iff

$$F(z) = \int_{\partial\mathbb{D}} P(e^{i\eta}, z)^\alpha |(\varphi^{-1})'(\varphi(e^{i\eta}))|^\alpha \frac{d\nu\circ\varphi}{d\nu} d\nu(e^{i\eta}).$$

By the uniqueness of the Karpelevich Representation, this holds if and only if $\frac{d\nu\circ\varphi}{d\nu} = |(\varphi^{-1})' \circ \varphi|^{-\alpha} = |\varphi'|^\alpha$. So F is Γ-invariant iff ν is Γ-conformal. □

Remark: The condition $\alpha \geq \frac{1}{2}$ is needed to be certain that the Karpelevich representation is unique, see [Ba].

Exercise: Fix $\varphi \in \text{Möb}(\mathbb{D})$ and suppose $F(z) := \int_{\partial\mathbb{D}} P(e^{i\theta}, z)^\alpha d\nu(e^{i\theta})$, where $\frac{d\nu\circ\varphi}{d\nu} = \lambda|\varphi'|^\alpha$. Show that $F(z) := \int_{\partial\mathbb{D}} P(e^{i\theta}, z)^\alpha d\nu(e^{i\theta})$ satisfies $F \circ \varphi = \lambda F$.

1.2.3 Babillot's construction

1. Start from a positive eigenfunction $F(\Gamma z)$ on $M = \Gamma \setminus \mathbb{D}$
2. Represent $\widetilde{F}(z) := F(\Gamma z) = \int_{\partial\mathbb{D}} P(e^{i\theta}, z)^\alpha d\nu(e^{i\theta})$ with $\alpha \geq \frac{1}{2}$, $\alpha(\alpha - 1) = \lambda$. Since F is Γ-invariant, ν is Γ-conformal with index α.
3. Form $dm = e^{\alpha s} d\nu(e^{i\theta}) ds dt$ and restrict m to a fundamental domain of Γ. Identify the restriction with a measure on $T^1(\Gamma \setminus \mathbb{D})$.

Theorem 1.5 (Babillot). *If ν is a finite Γ conformal measure with index α, then*

$$dm = e^{\alpha s} d\nu(e^{i\theta}) ds dt \tag{1.4}$$

is a Γ–invariant h–invariant locally finite measure on $T^1\mathbb{D}$. Its restriction to a fundamental domain of Γ determines an h–invariant measure on $T^1(\Gamma \setminus \mathbb{D})$.

Proof. The measure m is h–invariant, because in *kan*–coordinates h acts by a translation on the left coordinate. The measure m is Γ–invariant, because in *kan*-coordinates the Γ-action is given by $(*)$, so $dm \circ \varphi = e^{\alpha(s - \log|\varphi'|)} \frac{d\nu\circ\varphi}{d\nu} d\nu ds dt = e^{\alpha s} d\nu ds dt = dm$. □

The following theorem relates the symmetries of an eigenfunction F to the symmetries of the e.i.r.m. m which Babillot associates to F. Before we state the theorem we mention a simple fact: let $N(\Gamma) := \{\varphi \in \text{Möb}(\mathbb{D}) : \varphi\Gamma\varphi^{-1} = \Gamma\}$, then $N(\Gamma)$ acts isometrically on $M = \Gamma \setminus \mathbb{D}$ by $\varphi(\Gamma z) := \Gamma\varphi(z)$.

Theorem 1.6 (Babillot). *Suppose $F(z)$ is a non-trivial ergodic positive eigenfunction on $\Gamma \setminus \mathbb{D}$, and m is the horocycle ergodic invariant Radon measure associated to F by the Babillot construction.*

1. If $\Delta F = \alpha(\alpha-1)F$ with $\alpha \geq \frac{1}{2}$, then $m \circ g^s = e^{(\alpha-1)s} m$ for all s.
2. F is harmonic iff m is g-invariant
3. If $\varphi \in N(\Gamma)$ then $F \circ \varphi = \lambda F$ iff $m \circ d\varphi = \lambda m$

Proof. Write $F(z) = \int_{\partial\mathbb{D}} P(e^{i\theta}, z)^\alpha d\nu(e^{i\theta})$ with $\alpha \geq \frac{1}{2}$, then m is the restriction of $e^{\alpha s} d\nu ds dt$ to a fundamental domain of Γ.

The geodesic flow acts by $g^s \cdot (e^{i\theta_0}, s_0, t_0) = (e^{i\theta_0}, s_0 + s, e^{-s} t_0)$ in *kan*-coordinates (theorem 1.1). So $m \circ g^s = e^{(\alpha-1)s} m$. In particular, if F is harmonic, then $\alpha = 1$ and m is invariant w.r.t. the geodesic flow.

The isometry $d\varphi$ acts by $\varphi \cdot (e^{i\theta_0}, s_0, t_0) = (\varphi(e^{i\theta_0}), s_0 - \log|\varphi'(e^{i\theta_0})|, t_0 + f(\theta_0, s_0))$ for some f whose particular form is irrelevant. Thus $m \circ d\varphi = \left(\frac{1}{|\varphi'|^\alpha} \frac{d\nu\circ\varphi}{d\nu}\right) m$.

Arguing as in the proof of theorem 1.4, we find that $F \circ \varphi = \lambda F$ translates to $\frac{d\nu\circ\varphi}{d\nu} = \lambda|\varphi'|^\alpha$. So it is equivalent to $m \circ d\varphi = \lambda m$. □

Babillot's construction shows that every positive eigenfunction of Δ on $\Gamma \setminus \mathbb{D}$ gives rise to an invariant Radon measure for the horocycle flow on $T^1(\Gamma \setminus \mathbb{D})$. She realized that if M is highly non-compact, then it may admit many positive eigefunctions — and therefore many invariant Radon measures. Two questions arise naturally:

(a) Do all ergodic invariant Radon measures arise this way?
(b) Do extremal eigenfunctions lead to ergodic measures?

We will show that for a large class of surfaces the answers are positive.

1.3 All invariant measures arise from positive eigenfunctions

1.3.1 Tame hyperbolic surfaces

We describe a large collection of hyperbolic surfaces, possibly of infinite genus, for which Babillot's construction provides *all* invariant Radon measures. The building blocks of these surfaces are hyperbolic surfaces with boundary called "pairs of pants", which we now describe.

Pairs of Pants (pop). A *pair of pants* (pop) is the identification space of the right-angle hyperbolic octagon depicted in figure 1.3.

This is a hyperbolic surface with boundary, and the boundary consists of three closed geodesics of lengths $\ell_1, \ell_2, \ell_3 \in [0, \infty)$ called the *boundary components*. The case $\ell = 0$ corresponds to a cusp.

The *seams* of a pop are the geodesic segments which connect its boundary components. Any triplet (ℓ_1, ℓ_2, ℓ_3) determines a unique pop up to isometry. The *norm* of a pop Y is defined to be the maximal length of a boundary component. We denote it by $\|Y\|$. Simple polygons have Euler characteristic $\chi = 1$, so applying the Gauss–Bonnet Theorem to the octagon which represents a pop in $\mathbb{D}$ we find that all pops have the same area $8 \cdot \frac{\pi}{2} - 2\pi\chi = 2\pi$. It follows that pops with large norms must have at least one short seam.

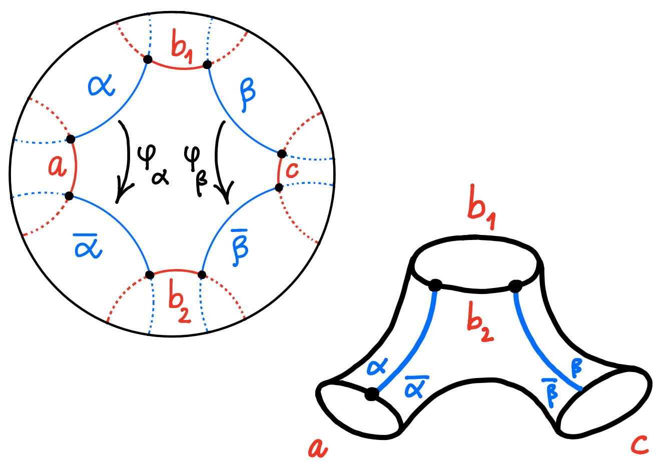

Fig. 1.3 A pair of pants

We can use pops to construct a variety of hyperbolic surfaces by gluing pops along boundary components of equal lengths. If we glue a finite or countable collection of pops in this way so that no boundary component remains "free," then the result is a hyperbolic surface (Figure 1.4). This surface does not have to be complete. But if all the pops we used have norm bounded above, then completeness is guaranteed, because it can be shown that the time it takes to cross a pop with norm less than t is bounded below by some $\varepsilon(t) > 0$, and therefore every geodesic ray can be continued indefinitely.

Many hyperbolic surfaces admit a countable decomposition into pops. To characterize them we need the notion of the *limit set* of Γ. Let Γ be a discrete subgroup of $\text{Möb}(\mathbb{D})$ without elements of finite order (we allow parabolic elements). The

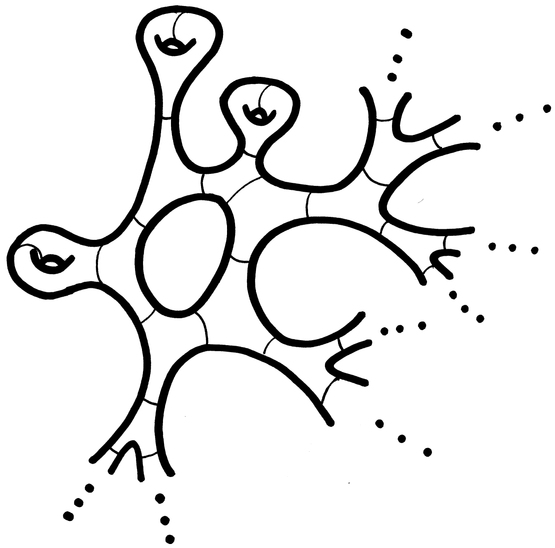

Fig. 1.4 A tame surface

limit set of Γ is the set $\Lambda(\Gamma)$ of accumulation points in the euclidean topology of $\{\varphi(z) : \varphi \in \Gamma\}$ for some (all) $z \in \mathbb{D}$. This set is independent of z.

Theorem 1.7. *$\Gamma \setminus \mathbb{D}$ has a decomposition into a finite or countable collection of pops iff $\Lambda(\Gamma) = \partial\mathbb{D}$.*

A proof can be found, for example, in [Hub]. Note that if Γ has parabolic elements, then $\Gamma \setminus \mathbb{D}$ has cusps, and some pops must have a boundary component with length zero and a seam with infinite length.

Tame surfaces. A connected orientable hyperbolic surface will be called *tame* if it can be divided into a finite or countable collection of pops Y_i such that the lengths of the boundary components are bounded away from infinity. If the boundary lengths are also uniformly bounded away from zero, then we call the surface *very tame*.

Every hyperbolic surface of finite area is tame, and every compact hyperbolic surface is very tame. It is easy to construct tame and very tame examples with infinite genus. For example, every regular cover of a compact hyperbolic surface is very tame, because it can be built from the pops of the compact surface it covers. Many more examples exist, see Figure 1.4.

1.3.2 The measure classification theorem for tame surfaces

The theorem is easiest to state for tame surfaces without cusps:

Theorem 1.8. *Suppose M is a tame hyperbolic surface without cusps, then*

1. *Every extremal positive eigenfunction leads to an ergodic horocycle invariant Radon measure through Babillot's construction.*
2. *Every non-trivial horocycle ergodic invariant Radon measure arises this way.*
3. *The mapping eigenfunctions→ measures is a bijection.*

Notice that all the measures m produced by Babillot's construction are automatically quasi-invariant for the geodesic flow. Indeed, $m \circ g^s = e^{(\alpha-1)s} m$ for all $s \in \mathbb{R}$, where $\alpha \geq \frac{1}{2}$ satisfies $\alpha(\alpha-1) = \lambda$, and λ is the eigenvalue.

Thus the theorem implies that for tame surfaces without cusps, all horocycle invariant Radon measures are quasi-invariant with respect to the geodesic flow.

This is not true in the presence of cusps. In this case additional measures appear, which sit on stable horocycles of vectors ω s.t. $g^t(\omega) \xrightarrow[t\to\infty]{}$cusp. Such measures satisfy $m \circ g^s \perp m$ for all $s \neq 0$. Henceforth we call horocycle invariant measures which sit on a single stable horocycle of a vector which tends to a cusp, *trivial measures*.

A positive eigenfunction $F(z)$ on $\Gamma \setminus \mathbb{D}$ is called *trivial* if it has the form $\sum c_k P(e^{i\theta_k}, z)^\alpha$ where $c_k \geq 0$. It can be shown that for tame surfaces, this is only possible when $e^{i\theta_k}$ are fixed points of parabolic elements of Γ.

Theorem 1.9. *Suppose M is a tame hyperbolic surface with cusps, then*

1. *every non-trivial extremal positive eigenfunction leads to a non-trivial ergodic horocycle invariant Radon measure through Babillot's construction.*
2. *every non-trivial horocycle e.i.r.m arises this way.*
3. *the mapping eigenfunctions→ measures is a bijection.*

So every ergodic horocycle invariant measure which is not supported on a single horocycle associated to a cusp, is quasi-invariant with respect to the geodesic flow.

Chapter 4 contains a sketch of the proof of part (2) of Theorem 1.8 in the special case of very tame surfaces.

1.3.3 Examples

Theorem (Furstenberg). *The horocycle flow on a connected compact hyperbolic surface is uniquely ergodic.*

Proof. Suppose m is a horocycle ergodic invariant Radon measure. By Theorem 1.8, m is the restriction of the measure $e^{\alpha s} d\nu ds dt$ to a compact fundamental domain. We claim that $\alpha = 1$ and $\nu = \text{const. Lebesgue}$.

We saw above that $m \circ g^s = e^{(\alpha-1)s} m$ for all s. Since M is compact, $m(T^1 M) < \infty$. Necessarily $\alpha = 1$ and the positive eigenfunction $F(z) = \int P(e^{i\theta}, z) d\nu(e^{i\theta})$ associated with m is harmonic. F lifts to bounded harmonic function on $\mathbb{D}$, which attains its maximum in a fundamental domain, whence in the interior of $\mathbb{D}$. By the maximum principle, F is constant. By the uniqueness of the Karpelevich representation, ν is equal to Lebesgue's measure up to a constant.

We have just shown that every horocycle ergodic invariant Radon measure is equal to the restriction of $e^s \frac{d\theta}{2\pi} dsdt$ to a fundamental domain. In particular, there is only one such measure. Since the volume measure is horocycle invariant, it must be the unique horocycle invariant probability measure. □

Theorem (Dani). *Suppose M is a connected hyperbolic surface with finite area. Every finite ergodic invariant measure is either supported on a single horocycle encircling a cusp, or is the uniform measure.*

Proof. Again, it is enough to show that every positive harmonic function on M is constant. Here is a sketch of the proof.

Let B_t denote the Brownian motion on M. Sullivan showed that the Brownian motion on a hyperbolic surface is recurrent iff the geodesic flow on the surface is ergodic. This is the case for hyperbolic surfaces of finite area. If F is a positive harmonic function on M, then it is a standard fact that $F(B_t)$ is a martingale, therefore $\lim_{t\to\infty} F(B_t)$ exists almost surely. Since B_t is recurrent, F must be constant. □

Theorem (Dani & Smillie). *Suppose M is a connected hyperbolic surface with finite area. Every ergodic invariant Radon measure is either supported on a single horocycle encircling a cusp, or is proportional to the volume measure.* [1]

Proof. This is a consequence of the fact that every non-constant extremal positive eigenfunction is trivial. A complete proof of this (well-known) fact can be found, e.g. in [LS2]. □

We now discuss some examples with infinite genus. A *co-compact periodic surface M* is a regular cover of a compact surface M_0. The group of deck transformations (or covering maps) of the cover is denoted by Cov, and satisfies $M/\mathrm{Cov} \cong M_0$.

The situation can be described algebraically as follows: $M = \Gamma \setminus \mathbb{D}$, $M_0 = \Gamma_0 \setminus \mathbb{D}$ where Γ_0 is a uniform lattice in $\text{Möb}(\mathbb{D})$ and $\Gamma \lhd \Gamma_0$. The covering map $p : M \to M_0$ is $p(\Gamma z) = \Gamma_0 z$, and every deck transformations is equal to a map of the form

$$D(\Gamma z) = \Gamma \varphi(z)$$

for some $\varphi \in \Gamma_0$. In particular, $\mathrm{Cov} \cong \Gamma_0/\Gamma$.

Every such surface is tame: it can be constructed from (possibly infinitely many) copies of the finite family of pops which make up the compact surface M_0.

[1] The theorem is a consequence of a stronger result by Dani & Smillie on the equidistribution of non-periodic horocycle orbits on hyperbolic surfaces of finite area. For more on equidistribution, see chapter 3.

Babillot-Ledrappier Measures: *Suppose M is a periodic surface with nilpotent covering group Cov without cusps. For every homomorphism $\varphi : Cov \to \mathbb{R}$ there is a horocycle e.i.r.m. s.t. $m_\varphi \circ D = e^{\varphi(D)} m_\varphi$ ($D \in Cov$). The measure m_φ is unique up to normalization. All ergodic invariant Radon measures arise this way.*

Complete details can be found in [LS2]. Here we limit ourselves to the case when Cov is *abelian* and just show that every e.i.r.m. m satisfies $m_\varphi \circ D = e^{\varphi(D)} m_\varphi$ for all $D \in \text{Cov}$ for some homomorphism $\varphi : \text{Cov} \to \mathbb{R}$.

By theorems 1.6 and 1.8 it is enough to show that every minimal and positive eigenfunction F satisfies $F \circ D = e^{\varphi(D)} F$ for all $D \in \text{Cov}$ and some homomorphism $\varphi : \text{Cov} \to \mathbb{R}$. The following proof of this fact uses ideas which can be traced to Margulis [M], Conze & Guivarc'h [CG], and Lyons & Sullivan [LS2].

We need the following fact: *For every $D \in Cov$ there is a constant $C(D)$ s.t.* $\text{dist}\,(p, D(p)) \le C(D)$ *for all* $p \in M$.

Here is the proof. Let $\widetilde{M}_0$ be a fundamental domain for the action of Cov on M. $\widetilde{M}_0$ lifts to a fundamental domain for the action of Γ_0 on $\mathbb{D}$, therefore $\widetilde{M}_0$ is pre-compact. For every $p \in M$ there is $p_0 \in \widetilde{M}_0$ and $D_0 \in \text{Cov}$ s.t. $p = D_0(p_0)$. Since Cov is abelian,

$$\begin{aligned}\text{dist}\,(p, D(p)) &= \text{dist}\,(D_0(p_0), (D \circ D_0)(p_0)) = \text{dist}\,(D_0(p_0), (D_0 \circ D)(p_0)) \\ &= \text{dist}\,(p_0, D(p_0)) \le C(D) := \sup\{\text{dist}\,(p_0, D(p_0)) : p_0 \in \widetilde{M}_0\}.\end{aligned}$$

$C(D)$ is finite because $\widetilde{M}_0$ is pre-compact and D is continuous.

We use this to show that $F \circ D \le \text{const.} F$. Fix p and find and isometry $M \cong \Gamma \setminus \mathbb{D}$ which maps p to the orbit of the origin o, Γo. Abusing notation we identify F with a function on $\Gamma \setminus \mathbb{D}$.

Let $\widetilde{F}(z) := F(\Gamma z)$ and let $\widetilde{F}(z) = \int_{\partial\mathbb{D}} P(e^{i\theta}, z)^\alpha d\nu(e^{i\theta})$ be the Karpelevich representation of $\widetilde{F}$. Let

$$K(D) := \max\{|P(e^{i\theta}, \xi)|^\alpha : \theta \in \mathbb{R}, \xi \in \mathbb{D}, \text{dist}\,(0, \xi) \le C(D)\} \le \left(\frac{1+\rho}{1-\rho}\right)^\alpha$$

for the $0 < \rho < 1$ s.t. $\text{dist}\,(0, \rho) = C(D)$ (dist :=hyperbolic distance).

Since p is represented by Γo, $D(p)$ is represented by $\Gamma \xi$ for some $\xi \in \mathbb{D}$ s.t. $\text{dist}\,(o, \xi) \le C(D)$. So

$$F(D(p)) = \int_{\partial\mathbb{D}} P(e^{i\theta}, \xi)^\alpha d\nu(e^{i\theta}) \le K(D)\nu(\partial\mathbb{D}) = K(D)\widetilde{F}(0) = K(D)F(p).$$

We see that $F(D(p)) \le K(D)F(p)$ for a constant $K(D)$ which is independent of p. So $F \circ D \le \text{const.} F$.

By minimality $F \circ D = c(D)F$ for some constant $c(D)$. What works for one D works for all D, and we must have $c(D_1D_2) = c(D_1)c(D_2)$. So $c(D) = \exp\varphi(D)$ for some additive homomorphism φ. □

1.4 Open problem

Suppose M is a connected, orientable, complete, hyperbolic surface *which is not tame*. Is it true that then every e.i.r.m which does not arise from Babillot's construction is necessarily carried by a single horocycle?

1.5 Notes and references

For a good concise introduction to hyperbolic geometry, see Beardon's chapter in [Se1], and for a nice account of the Killing-Hopf Theorem on the representation of a hyperbolic surface as an orbit space, see [Sti]. Karpelevich's Theorem is proved in [Kar] and [Gui]. The theory of conformal measures (or "densities") for Fuchsian groups is developed in the papers of Patterson [Pat] and Sullivan [Su1, Su2], see also [Ro2]. for a A complete treatment of the positive eigenfunctions on nilpotent covers of compact surfaces can be found in [LP].

The horocycle flow first appeared in the works of E. Hopf and Hedlund on the ergodicity of the geodesic flow. Hedlund, who later studied the flow for its own sake, clarified many of its dynamical properties.

The horocycle invariant measures were first classified for compact surfaces by Furstenberg [Fu1]. The case of non-compact surfaces with finite area was done by Dani [Da] for finite invariant measures. Later work by Dani & Smillie [DS] shows the equidistribution of all non-periodic horocycles, which implies in particular that there are no infinite invariant Radon measures in this case. Ratner's theory [Ra2] classifies the finite measures for general unipotent flows (of which the horocycle flow is a particular case). It implies that the horocycle flow on a hyperbolic surface of infinite area has no finite invariant measures at all, except on closed horocycles encircling cusps. *Thus in infinite area, all non-trivial e.i.r.m. are infinite.*

Burger constructed a non-trivial infinite ergodic invariant Radon measure for the horocycle flow on a hyperbolic surface of infinite area and finite genus [Bu], and showed under additional assumptions which were later removed by Roblin [Ro2] that this is the unique e.i.r.m. (up to scaling) not carried by a single orbit (this is not the volume measure: that measure is not ergodic).

Babillot & Ledrappier were the first to discover that in infinite genus, there could be many globally supported e.i.r.m's [BL2]: for $\mathbb{Z}^d$–covers of compact hyperbolic surfaces, they exhibited a d-parameter family of mutually singular globally supported e.i.r.m's ("Babillot-Ledrappier measures"). The original construction used symbolic dynamics and thermodynamic formalism. Later Babillot made the connection to positive eigenfunctions of the Laplacian, introduced what we called above "Babillot's construction", and used it to construct e.i.r.m's for nilpotent covers of compact surfaces [Ba].

The question whether all e.i.r.m's arise this way was settled positively in [Sa1] for $\mathbb{Z}^d$-covers, then in [LS2] for all regular covers of hyperbolic surfaces of finite area, then in [Sa2] for a class of surfaces which contains all tame hyperbolic surfaces.

Some of the results of this section extend to variable negative curvature, and to higher-dimension, see [Sa1], [L], [Po].

Chapter 2
Ergodic properties of horocycle invariant Radon measures

We review some of the basics of infinite ergodic theory, and apply them to horocycle flows on $\mathbb{Z}^d$–covers of compact hyperbolic surfaces.

2.1 Infinite ergodic theory

2.1.1 Three basic facts

Let Ω be a complete separable metric space, let $\mathscr{F}$ be its collection of Borel sets, and fix some σ-finite Borel measure μ.

A *measure preserving map* on $(\Omega, \mathscr{F}, \mu)$ is a measurable map $T : \Omega \to \Omega$ s.t. $\mu(T^{-1}E) = \mu(E)$ for all $E \in \mathscr{F}$. T is called *ergodic* if for every set $E \in \mathscr{F}$ s.t. $T^{-1}E = E$, either E has measure zero or $\Omega \setminus E$ has measure zero. T is called *conservative* if every measurable set W s.t. $\{T^{-n}W\}_{n\geq 0}$ are pairwise disjoint, $\mu(W) = 0$.

Theorem 2.1 (Halmos). *If a measure preserving map is ergodic and conservative, then for every $0 \leq f \in L^1$ s.t. $\int f > 0$, $\sum_{n\geq 0} f \circ T^n = \infty$ μ-almost everywhere.*

So if $\mu(E) > 0$, then a.e. every orbit visits E infinitely many times. Contrast this with the behavior of the map $x \mapsto x+1$ on $\mathbb{Z}$ which is ergodic, but not conservative.

A *measure preserving flow* on $(\Omega, \mathscr{F}, \mu)$ is a one-parameter family of measure preserving maps $T^t : \Omega \to \Omega$ ($t \in \mathbb{R}$) s.t. $(t,x) \mapsto T^t(x)$ is measurable, and $T^{t+s} = T^t \circ T^s$ for all $s,t \in \mathbb{R}$.

A flow is called *conservative* if its time one map is conservative. It can be shown that in this case, if $\mu\{x : T^t(x) = x\} = 0$ for all $t \neq 0$, then $T^t : \Omega \to \Omega$ are conservative for all $t \neq 0$ [A, §1.6].

A measurable function $f : \Omega \to \mathbb{R}$ is called *invariant* if for a.e. x, $f(T^t x) = f(x)$ for all $t \in \mathbb{R}$ (notice the order of the quantifiers). A flow is called *ergodic* if every invariant function is constant almost everywhere.

Theorem 2.2 (Hopf's Ratio Ergodic Theorem). *Let $\{T^t\}$ be a conservative measure preserving flow on $(\Omega, \mathscr{F}, \mu)$. For every $f, g \in L^1$ s.t. $\int g d\mu > 0$, the limit*

$$\lim_{T\to\infty} \frac{\int_0^T f\circ T^t\,dt}{\int_0^T g\circ T^t\,dt}$$

exists a.e. and is an invariant function. If $\{T^t\}$ *is ergodic, the limit is* $\int f d\mu / \int g d\mu$ μ*–almost everywhere.*

Corollary 2.1. *Let* $\{T^t\}$ *be an ergodic conservative measure preserving flow on* $(\Omega, \mathscr{F}, \mu)$*. Suppose* $\mu(\Omega) = \infty$*, then for every* $f \in L^1$ *s.t.* $\int f d\mu > 0$*,*

$$\int_0^\infty f(T^t(\omega))dt = \infty, \quad \text{and} \quad \frac{1}{T}\int_0^T f(T^t(\omega))dt \xrightarrow[T\to\infty]{} 0 \text{ a.e. } \omega \in \Omega.$$

Proof. The divergence of $\int_0^\infty f\circ T^t dt$ is because of the conservativity of T^1 and the decomposition $\int_0^\infty f\circ T^t dt = \sum_{n=0}^\infty F\circ T^n$ where $F = \int_0^1 f\circ T^t dt$. The convergence of $\frac{1}{T}\int_0^T f\circ T^t dt$ to zero follows from the ratio ergodic theorem by taking $g =$indicator of an arbitrarily large set of finite measure. ☐

The corollary deserves some reflection. Suppose A is a measurable set with finite measure and $f = 1_A$ (the indicator of A, equal to one on A and to zero outside A), then $\frac{1}{T}\int_0^T f(T^t(\omega))dt$ is the frequency of times $0 < t < T$ that $T^t(\omega) \in A$. The previous theorem says that a.e. orbit visits A an infinite amount of time, but with asymptotic frequency zero. Still we can say that some sets are visited more often than others, because by the ratio ergodic theorem, if $\mu(A) > \mu(B)$ then A will be visited more often than B by a factor $\rho = \mu(A)/\mu(B)$.

It is natural to ask if there is another sequence of normalizing constants $a_T = o(T)$ s.t. $\frac{1}{a_T}\int_0^T f(T^t(\omega))dt \xrightarrow[T\to\infty]{} \int f d\mu$ a.e. for $f \in L^1$. The answer is a resounding "No:"

Theorem 2.3 (Aaronson). *Let* $\{T^t\}$ *be an ergodic conservative measure preserving flow on* $(\Omega, \mathscr{F}, \mu)$*, and let* $f \in L^1$ *be a non-negative function with non-zero integral. If* $\mu(\Omega) = \infty$*, then for every* $a_T > 0$*, either* $\liminf \frac{1}{a_T}\int_0^T f(T^t(\omega))dt = 0$ *a.e., or* $\limsup \frac{1}{a_T}\int_0^T f(T^t(\omega))dt = \infty$ *a.e.*

Thus $\int_0^T f(T^t(\omega))dt \neq [1+o(1)]a_T \int f d\mu$ a.e. as $T\to\infty$ for any normalization a_T.

2.1.2 High order ergodic theorems

Sometimes, it is possible to find normalizing constants a_T s.t. $\int_0^T f(T^t(x))dt \approx a_T \int f d\mu$ a.e. in some weaker sense than asymptotic equivalence. The precise meaning of $\approx$ varies from case to case. The best one can hope for is

$$\int_0^T f(T^t(x))dt = [1+o(1)]\cdot a_T \cdot \left(\int f d\mu\right)\cdot \mathrm{Osc}_T(x), \text{ a.e. as } T\to\infty, \qquad (2.1)$$

where a_T is independent of f and x, and $\mathrm{Osc}_T(x)$ is a sequence of positive measurable functions which are independent of f, so that $\int_{\Omega_0} \mathrm{Osc}_T d\mu = 1$ for some fixed $\Omega_0 \in \mathscr{F}$ of finite positive measure. By Corollary 2.1, $a_T \to \infty$ and $a_T = o(T)$. By Aaronson's theorem, $\mathrm{Osc}_T(x)$ oscillates a.e. without converging. The ratio ergodic theorem is reflected by the independence of $\mathrm{Osc}_T(x)$ from f.

Suppose we managed to prove something like (2.1), and suppose we have sufficient information on the almost sure behavior of $\mathrm{Osc}_T(x)$ to design a summability method (S) which regularizes its oscillations, so that $\mathrm{Osc}_N(x) \xrightarrow[n\to\infty]{(S)} 1$ a.e. Then the result would be what A. Fisher calls a *"high order ergodic theorem:"* $\frac{1}{a_T}\int_0^T f(T^t(x)) \xrightarrow[T\to\infty]{(S)} \int f d\mu$ a.e. We will see examples below.

2.1.3 Squashability

Some ergodic conservative measures do not admit any high-order ergodic theorems. To make the statement as strong as possible, we will use a definition of a "generalized ergodic theorem" which is as weak as possible, and produce an obstruction.

The essence of the ergodic theorem is that it allows to determine the measure of a set E from the set of times a typical orbit visits E. Here is an abstraction of this idea. Given an infinite measure preserving flow $\mathsf{T} = (\Omega, \mathscr{F}, \mu, \{T^t\})$, a measurable set E of finite positive measure, and $\omega \in \Omega$, let

$$x_{E,\omega}(t) := 1_E(T^t(\omega)) = \begin{cases} 1 & T^t(\omega) \in E \\ 0 & \text{otherwise} \end{cases} \quad (t \geq 0).$$

The following definition is due to Jon Aaronson.

Generalized Law of Large Numbers: *A generalized law of large numbers (GLLN) for T is a function $L : \{0,1\}^{\mathbb{R}_+} \to [0,\infty)$, $L = L[x(\cdot)]$ such that for every $E \in \mathscr{F}$ of finite measure, $L[x_{E,\omega}(\cdot)] = \mu(E)$ for a.e. ω.*

For example, if T satisfies the high order ergodic theorem with summability method (S), $\frac{1}{a_T}\int_0^T f(T^t x)dt \xrightarrow[T\to\infty]{(S)} \int f d\mu$ a.e., then the following is a GLLN for T:

$$L[x(\cdot)] := \begin{cases} \underset{T\to\infty}{\text{S-lim}} \frac{1}{a_T}\int_0^T x(t)dt & \text{when the integral and the (S)-limit exist,} \\ -666 & \text{otherwise.} \end{cases}$$

Here "S-lim" denotes the limit according to (S).

Squashibility: *T is called* squashable*, if there exists a measurable map $Q : \Omega \to \Omega$ s.t. $Q \circ T = T \circ Q$ and $\mu \circ Q^{-1} = c\mu$ with $c \neq 0, 1$.*

This pathology can only exist in infinite measure, as can be seen by evaluating both sides of the equation $\mu \circ Q^{-1} = c\mu$ on Ω.

Theorem 2.4 (Aaronson). *A squashable infinite measure preserving system has no generalized laws of large numbers.*

Proof. [A] Suppose $Q \circ T = T \circ Q$ and $\mu \circ Q^{-1} = c\mu$ where $c \neq 0, 1$. Let L be a GLLN. Since $\mu \circ Q^{-1} = c\mu$ with $c \neq 0$, what holds a.e. for ω, also holds a.e. for $Q(\omega)$. So $\mu(E) \overset{\text{a.e.}}{=} L[x_{E,Q(\omega)}(\cdot)] = L[x_{Q^{-1}E,\omega}(\cdot)] \overset{\text{a.e.}}{=} \mu(Q^{-1}E) = c\mu(E)$, and $c = 1$. But $\mu \circ Q^{-1} = c\mu$ with $c \neq 1$ by assumption. □

2.1.4 Ergodicity of the volume measure

The volume measure on the unit tangent bundle of a hyperbolic surface is always invariant under the geodesic and horocycle flows, but in infinite area , it is not always conservative or ergodic.

The following two theorems characterize the surfaces where it is conservative and ergodic. Let Γ be a discrete subgroup of Möb$(\mathbb{D})$ and let $M := \Gamma \setminus \mathbb{D}$.

Theorem 2.5 (Hopf-Tsuji-Sullivan). *The following are equivalent:*

1. *The geodesic flow on T^1M is conservative with respect to the volume measure.*
2. *The geodesic flow on T^1M is ergodic with respect to the volume measure.*
3. *The Brownian motion on M is recurrent.*
4. *The Poincaré series $\sum_{\gamma \in \Gamma} e^{-sd(0,\gamma(0))}$ diverges at $s = 1$.*

Theorem 2.6 (Kaimanovich). *The horocycle flow on T^1M is ergodic with respect to the volume measure iff all bounded harmonic functions on M are constant. In this case the horocycle flow is also conservative.*

If M has finite hyperbolic area, then the geodesic flow and the horocycle flow are both ergodic [Ho1, Ho2]. Hopf's argument has the merit of extending to variable negative curvature (even to Anosov flows). But in constant curvature, there are shorter modern proofs, see [Bek].

If M has infinite hyperbolic area but finite genus, then it must have funnels (Figure 3.1.2). The volume measure ceases to be ergodic or conservative, because a.e. geodesic and a positive volume of horocycles escape through a funnel. But there are other natural ergodic conservative invariant measures on the non-wandering set of these flows, see [Su1, Ro2, Bu, Scha].

In infinite genus, the picture is mixed. For example, in the case of $\mathbb{Z}^d$-covers of compact hyperbolic surfaces, Mary Rees proved in [Ree] that the geodesic flow is ergodic when $d \leq 2$ and non-ergodic when $d \geq 3$. But the horocycle flow is ergodic and conservative for all $d \geq 1$ [BL2, Sol, Po, Cou].

2.2 Example: $\mathbb{Z}^d$-covers of compact hyperbolic surfaces

We demonstrate these ideas in the special case of the horocycle flow on a regular $\mathbb{Z}^d$-cover M of a compact hyperbolic surface M_0 (cf. §1.3.3). Let Cov $\cong \mathbb{Z}^d$ the group of covering maps ("deck transformations"). These are isometries of M and as such, they commute with the geodesic and horocycle flows.

In the last chapter we saw that for every homomorphism $\psi : \mathrm{Cov} \to \mathbb{R}$ there exists an ergodic invariant Radon measure (e.i.r.m) m_ψ s.t.

$$m_\psi \circ dD = e^{\psi(D)} m_\psi \text{ for all } D \in \mathrm{Cov}.$$

This measure is uniquely determined up to scaling, and all e.i.r.m's take this form.

If $\psi \not\equiv 0$ then there exists a deck transformation D s.t. $\psi(D) \neq 0$, and then $m_\psi \circ dD = e^{\psi(D)} m_\psi = c m_\psi$ where $c \neq 0, 1$. So m_ψ is squashable, and does not admit high-order ergodic theorems or other generalized laws of large numbers.

If $\psi \equiv 0$ then $m_0 := m_\psi$ is proportional to the volume measure (exercise). It turns out that this measure does indeed satisfy a high-order ergodic theorem. Let (S) denote the Cesàro summability method with weights $1/T \ln T$, i.e.

$$A(T) \xrightarrow[T\to\infty]{(S)} A \quad \text{iff} \quad \frac{1}{\ln\ln N} \int_3^N \frac{A(T)dT}{T \ln T} \xrightarrow[N\to\infty]{} A \tag{S}$$

$(T \ln T)^{-1}$ are the "weights" and $1/\ln\ln N \sim \int_3^N \frac{dT}{T \ln T}$ is the "sum of the weights." The domain of integration starts at 3 because $\ln\ln T = 0$ for $T = e$.

Theorem 2.7 (Ledrappier-S.). *Let $a(T) := T/(\ln T)^{d/2}$, then for every $f \in L^1(m_0)$ $\frac{1}{a(T)} \int_0^T f(h^t(\omega))dt \xrightarrow[T\to\infty]{(S)}$ const. $\int f dm_0$. The constant is non-zero.*

Note that $a(T)$ depends on the surface through d.

We see that the horocycle flow is "conditionally uniquely ergodic" in the following sense: up to scaling, it has just one e.i.r.m. with a GLLN (the volume measure). All other (infinitely many) e.i.r.m's are squashable.

The proof of theorem 2.7 relies on an analysis of the oscillatory behavior of $\frac{1}{a(T)} \int_0^T f(h^t(\omega))dt$. It turns out that the oscillations are driven by a "random walk" that the *geodesic* flow performs on M.

To describe this walk, enumerate $\mathrm{Cov} = \{D_{\underline{a}} : \underline{a} \in \mathbb{Z}^d\}$ in such a way that $D_{\underline{a}} \circ D_{\underline{b}} = D_{\underline{a}+\underline{b}}$. Since M is a $\mathbb{Z}^d$-cover of a compact surface M_0, there is a precompact connected lift $\widetilde{M}_0$ of M_0 to M such that $M = \biguplus_{\underline{a}\in\mathbb{Z}^d} M_{\underline{a}}$ where $M_{\underline{a}} = D_{\underline{a}}(\widetilde{M}_0)$. The function $\xi : T^1M \to \mathbb{Z}^d$, $\xi(\omega) = \underline{a}$ on $\widetilde{M}_{\underline{a}}$ is called a *$\mathbb{Z}^d$–coordinate* on M. It extends in an obvious way to T^1M.

Let g^s denote the geodesic flow on T^1M. If we choose ω at random at $\widetilde{M}_0$, then $\xi(g^s(\omega))$ becomes a stochastic process taking values in $\mathbb{Z}^d$. This is the *random walk performed by the geodesic flow on M.*

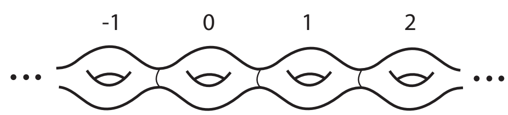

Fig. 2.1 A $\mathbb{Z}$-coordinate on a $\mathbb{Z}$-cover of a surface of genus two

Theorem 2.8 (Ratner, Katsuda & Sunada). *If ω is chosen randomly uniformly in $\widetilde{M}_0$, then $\xi(g^s(\omega))/\sqrt{s} \xrightarrow[s\to\infty]{dist} \mathbf{N}$ where $\mathbf{N}$ is a Gaussian random variable on $\mathbb{R}^d$ with mean zero and positive definite covariance matrix $N := (\mathbb{E}(N_i N_j))_{d\times d}$.*

Let $\|\cdot\|_N$ denote the norm on $\mathbb{R}^d$, $\|\mathbf{v}\|_N := \sqrt{\mathbf{v}^t N \mathbf{v}}$, and let $\sigma := \sqrt[d]{|\det(N)|}$. The notation $a = e^{\pm\varepsilon} b$ is shorthand for $e^{-\varepsilon} \le \frac{a}{b} \le e^{\varepsilon}$.

Theorem 2.9 (Ledrappier & S.). *Suppose $0 \le f \in L^1$ and $\frac{1}{m_0(\widetilde{M}_0)} \int f dm_0 = 1$. For every $\varepsilon > 0$, for m_0–almost every ω there is $T_0 > 0$ s.t.*

$$\int_0^T f(h^t(\omega))dt = e^{\pm\varepsilon} \cdot \left(\frac{1}{(2\pi\sigma)^{d/2}} \frac{T}{(\ln T)^{d/2}} \right) \cdot \exp\left(-\frac{1}{2}(1 \mp \varepsilon) \left\| \frac{\xi(g^{\ln T}(\omega))}{\sqrt{\ln T}} \right\|_N^2 \right).$$

This is in the spirit of (2.1): $\exp\left(-\frac{1}{2}(1\mp\varepsilon)\left\|\frac{\xi(g^{\ln T}(\omega))}{\sqrt{\ln T}}\right\|_N^2\right)$ are upper and lower bounds for the "oscillating term" and are themselves oscillating. These bounds show that the oscillating term converges in distribution to $\exp(-\frac{1}{2}\|\mathbf{N}\|_N^2)$, a random variable with positive variance, indicating highly oscillatory behavior.

The speed of the oscillations is very slow: They happen on time scale $\ln T$. This means that $\frac{1}{a(T)} \int_0^T f(h^t(\omega))dt$ will deviate significantly from $\frac{1}{m_0(\widetilde{M}_0)} \int f dm_0$ on time intervals of exponential size. These are the time intervals when $\frac{\xi(g^{\ln T}(\omega))}{\sqrt{\ln T}}$ is far from its mean, zero. There will be infinitely many such intervals.

A more delicate analysis of the statistical behavior of $\xi(g^s(\omega))/\sqrt{s}$ reveals that

$$\lim_{N\to\infty} \frac{1}{\ln\ln N} \int_3^N \exp\left(-\frac{1}{2} \left\| \frac{\xi(g^{\ln T}(\omega))}{\sqrt{\ln T}} \right\|_N^2 \right) \frac{dT}{T \ln T} = \sqrt{2} \text{ almost everywhere.}$$

This why the summability method (S) is used in theorem 2.7. For details, and the proof of Theorem 2.9, see [LS1].

2.3 Open problem

Find high-order ergodic theorems for horocycle flows on regular covers of compact hyperbolic surfaces with *nilpotent* groups of deck transformations.

2.4 Notes and references

The main reference for §2.1 is Aaronson's book [A]. Theorems 2.1, 2.2 and 2.3 follow from Proposition 1.2.2, Theorem 2.2.5, and Theorem 2.4.2 there, except that [A] states these results for maps instead of flows.[1]

For a proof of the Hopf-Tsuji-Sullivan theorem, and more information on the ergodic theoretic properties of the geodesic flow in case of infinite volume, see [Su2], [A] and [Ro2]. For proof of Kaimanovich's theorem and many illuminating examples, see [Kai]. That paper and [Sta] contain a discussion of other dynamical properties of these flows.

Examples of high order ergodic theorems can be found in [ADF] and [Fi2], see also [Fi1]. Examples of other generalized laws of large numbers can be found in the book of Aaronson [A], who was the one to introduce them (under the name "laws of large numbers"). The phenomenon of different measures for the same dynamical systems, some with a GLLN and others without appears for non-compact skew-products, see e.g. [AW], [ANSS].

The main reference for §2.2 is [LS1], which contains the proofs of Theorems 2.7 and 2.9. The central limit theorem in theorem 2.8 follows from [Ra1]. The positive definiteness of the covariance matrix N is shown in [KS]. High order ergodic theorems for $\mathbb{Z}^d$–covers of non-compact hyperbolic surfaces with finite area are given in [LS3]. The normalization constants $a(T)$ and the oscillating term are different in this case, because of the effect of the cusps.

---

[1] Here is how to deduce the ratio ergodic theorem for flows from the ratio ergodic theorem for maps [Ho1]: It is enough to treat the case $0 \le f \in L^1$, $0 \le g \in L^\infty \setminus \{0\}$, because if we know the ratio ergodic theorem for all $f \in L^1_+, g \in L^\infty_+$ then we know it for all $f \in L^1, g \in L^1_+$. Let $S := T^1$, $F(x) := \int_0^1 f(T^t x)dt$, and $G(x) := \int_0^1 g(T^t x)dt$.

$$\left| \frac{\int_0^\tau f \circ T^t dt}{\int_0^\tau g \circ T^t dt} - \frac{\sum_{k=0}^{\lfloor \tau \rfloor} F \circ S^k}{\sum_{k=0}^{\lfloor \tau \rfloor} G \circ S^k} \right| \le \frac{F \circ S^{\lfloor \tau \rfloor}}{\sum_{k=0}^{\lfloor \tau \rfloor - 1} G \circ S^k} + \frac{G \circ S^{\lfloor \tau \rfloor}}{\sum_{k=0}^{\lfloor \tau \rfloor - 1} G \circ S^k}$$
$$= \frac{\sum_{k=0}^{\lfloor \tau \rfloor} F \circ S^k}{\sum_{k=0}^{\lfloor \tau \rfloor} G \circ S^k} \cdot \frac{\sum_{k=0}^{\lfloor \tau \rfloor} G \circ S^k}{\sum_{k=0}^{\lfloor \tau \rfloor - 1} G \circ S^k} - \frac{\sum_{k=0}^{\lfloor \tau \rfloor - 1} F \circ S^k}{\sum_{k=0}^{\lfloor \tau \rfloor - 1} G \circ S^k} + \frac{G \circ S^{\lfloor \tau \rfloor}}{\sum_{k=0}^{\lfloor \tau \rfloor - 1} G \circ S^k} \underset{\tau \infty}{\longrightarrow} 0 \text{ a.e.}$$

by the ratio ergodic theorem for S, and since $\|G\|_\infty < \infty$ and $\sum G \circ S^k = \infty$ (conservativity). Thus the existence of $\lim \frac{\sum_{k=0}^{\lfloor \tau \rfloor} F \circ S^k}{\sum_{k=0}^{\lfloor \tau \rfloor} G \circ S^k}$ (ratio ergodic theorem for maps) implies the existence of $\lim \frac{\int_0^\tau f \circ T^t dt}{\int_0^\tau g \circ T^t dt}$ (ratio ergodic theorem for flows).

Chapter 3
Generic points and equidistribution

This chapter describes the points with equidistributed horocycles in the case of $\mathbb{Z}^d$-covers of compact surfaces.

3.1 Generic points

3.1.1 Definition

Let $\varphi^t : X \to X$ denote a continuous flow on a second countable locally compact metric space X. Let $C_c(X) := \{f : X \to \mathbb{R} : f \text{ is continuous, with compact support}\}$. Recall that a Borel measure μ is a Radon measure iff every $f \in C_c(X)$ is absolutely integrable w.r.t. μ.

Theorem 3.1. *Suppose μ is a conservative ergodic invariant Radon measure on a second countable locally compact metric space X, then for a.e. $x \in X$*

$$\frac{\int_0^T f(\varphi^t(x))dt}{\int_0^T g(\varphi^t(x))dt} \xrightarrow[T\to\infty]{} \frac{\int f d\mu}{\int g d\mu} \quad \textit{for all } f,g \in C_c(X) \textit{ s.t. } \int g d\mu > 0. \tag{3.1}$$

Proof. By the topological assumptions on X, there exists a countable family of functions $\mathscr{F} \subset C_c(X)$ s.t. for every $0 \le f \in C_c(X)$, for every $\varepsilon > 0$, there are $u, v \in \mathscr{F}$ s.t. $u \le f \le v$ and $\|u - v\|_1 < \varepsilon$.

By the ratio ergodic theorem, for fixed $u, v \in \mathscr{F}$ s.t. $\int v d\mu > 0$,

$$X(u,v) := \left\{ x : \int_0^T u(\varphi^t(x))dt \Big/ \int_0^T v(\varphi^t(x))dt \xrightarrow[T\to\infty]{} \int u d\mu \Big/ \int v d\mu \right\}$$

has full measure. Fix $v \in \mathscr{F}$ s.t. $\int v d\mu > 0$. Since $\mathscr{F}$ is countable, $X_0 := \bigcap_{u \in \mathscr{F}} X(u,v)$ has full measure. Since every non-negative $f \in C_c(X)$ can be sandwiched between $\mathscr{F}$–elements, every point in X_0 satisfies (3.1) with $g = v$, whence with general g. $\square$

Any point which satisfies (3.1) is called *generic* (for μ). If X is a manifold and μ is the volume measure, then we say that the orbit of the point is *equidistributed*.

3.1.2 Horocycle flows on surfaces with finite genus

Let M be a complete, connected, orientable, hyperbolic surface with unit tangent bundle T^1M, geodesic flow g^s, and horocycle flow h^t. The generic points are known in the following cases.

Compact surfaces (Furstenberg): Every $\mathbf{v} \in T^1M$ is equidistributed.

Non-compact surfaces with finite area (Dani & Smillie): Such surfaces have cusps. Vectors $\mathbf{v}$ s.t. $g^t(\mathbf{v}) \xrightarrow[T\to\infty]{}$cusp have periodic horocycles, and are obviously generic for the normalized Lebesgue measure on their horocycles. All other vectors are equidistributed.

Surfaces with infinite area and finite genus (Schapira): Such surfaces have funnels, and funnels contain non-recurrent horocycles (Figure 3.1.2). More precisely, call $\mathbf{v} \in T^1M$ *exceptional* if $\int_0^\infty \delta_{h^t(\mathbf{v})}dt$ is a Radon measure and *very exceptional* if $\int_{-\infty}^\infty \delta_{h^t(\mathbf{v})}dt$ is a Radon measure. Such vectors are not generic for any measure.

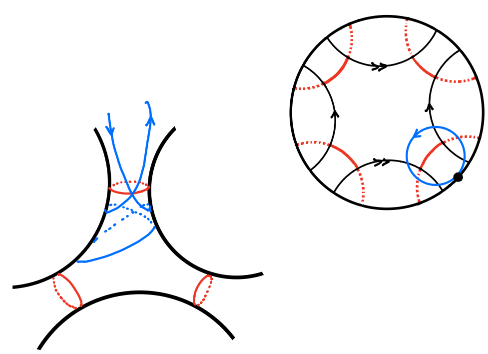

Fig. 3.1 A surface with a funnel and a very exceptional horocycle

It can be shown that with respect to the volume measure, almost every orbit is exceptional. Thus, although the volume measure is invariant and Radon, it is not ergodic, and its ergodic components are not conservative.

There is however another ergodic invariant Radon measure, called the *Burger-Roblin measure*, which sits on the non-wandering set of the horocycle flow. Barbara Schapira showed that every unit tangent vector is either exceptional, periodic (and generic for the unique invariant probability measure on its orbit), or generic for the Burger-Roblin measure. (The first result of this type, which dealt with the equidistribution of symmetric ergodic sums $\int_{-T}^{T}$, and which assumed more on the surface, is due to Marc Burger.)

The Burger-Roblin measure, the measures on periodic horocycles, and the infinite non-conservative measures sitting on horocycles of very exceptional vectors constitute a full list of the e.i.r.m's in this case (Roblin).

Ratner Theory: Ratner theory characterizes the generic points for all *finite* ergodic invariant measures for general unipotent flows, e.g. the horocycle flow.

3.1.3 $\mathbb{Z}^d$-covers of compact surfaces

Suppose M is a regular cover of a compact hyperbolic surface M_0, with covering group Cov $\cong \mathbb{Z}^d$. Parameterize this group by $\mathbb{Z}^d$:

$$\text{Cov} := \{D_{\underline{a}} : \underline{a} \in \mathbb{Z}^d\}\ ,\ D_{\underline{a}} \circ D_{\underline{b}} = D_{\underline{a}+\underline{b}}.$$

Recall that the e.i.r.m in this case are $\{cm_\varphi : c > 0, \varphi : \text{Cov} \to \mathbb{R} \text{ is a homomorphism}\}$ where $m_\varphi \circ D = e^{\varphi(D)} m_\varphi$ for all $D \in \text{Cov}$. The measure m_0, corresponding to the trivial homomorphism is the volume measure.

Let $\widetilde{M}_0$ denote a connected lift of M_0 to M, then $M = \biguplus_{\underline{a}\in\mathbb{Z}^d} D_{\underline{a}}(\widetilde{M}_0)$. Recall that the $\mathbb{Z}^d$*–coordinate* of M is the function $\xi : T^1M \to \mathbb{Z}^d$ s.t. $\xi(\mathbf{v}) = \underline{a}$ for vectors based in $D_{\underline{a}}(\widetilde{M}_0)$ (Figure 2.2). The *drift* of $\mathbf{v} \in T^1M$ is the following limit, if it exists:

$$\Xi(\mathbf{v}) = \lim_{s\to\infty} \frac{1}{s}\xi(g^s(\mathbf{v})).$$

(g^s =geodesic flow). Let $\mathfrak{C} := \overline{\text{convex hull}\{\Xi(\mathbf{v}) : \mathbf{v} \text{ s.t. } \Xi(\mathbf{v}) \text{ exists}\}}$.

Theorem 3.2 (Babillot & Ledrappier). *For every Babillot-Ledrappier measure m_φ there is a vector $\Xi_\varphi \in int(\mathfrak{C})$ s.t. $\Xi(\cdot) = \Xi_\varphi$ m_φ–almost everywhere. The map $m_\varphi \mapsto \Xi_\varphi$ is one-to-one. For the volume measure m_0, $\Xi_{\underline{0}} = \underline{0}$.*

Thus every e.i.r.m has constant drift a.e., and the value of this drift characterizes the measure up to scaling. The following theorem says that the drift also characterizes the generic vectors:

Theorem 3.3 (S. & Schapira). *A vector $\mathbf{v}$ is generic for some e.i.r.m. m iff $\Xi(\mathbf{v})$ exists and belongs to $int(\mathfrak{C})$. In this case $m = m_\varphi$ for the unique φ s.t. $\Xi_\varphi = \Xi(\mathbf{v})$. In particular, the horocycle of $\mathbf{v}$ is equidistributed iff $\Xi(\mathbf{v}) = \underline{0}$.*

3.2 Sketch of proof for the volume measure

The proof is too technical to include in these notes, so we just indicate its structure, and refer the reader to [SSc] for the details. Throughout this section, M is a $\mathbb{Z}^d$-cover of a compact hyperbolic surface, h is the horocycle flow, and g is the geodesic flow.

3.2.1 Sufficiency

The key is the following variation on theorem 2.9 which can be proved using harmonic analysis in the spirit of [BL1] and [La].

Lemma 3.1. *Fix $0 \leq f \in C_c(T^1M)$ s.t. $\int f dm_0 > 0$. For every $\varepsilon > 0$ there is $\delta > 0$ and $T_0 > 0$ s.t. for all $T > T_0$ and for every $\mathbf{v} \in T^1M$, if $\|\xi(g^{\ln T}(\mathbf{v}))/\ln T\| < \delta$, then*

$$\int_0^T f(h^t(\mathbf{v}))dt \leq e^{\varepsilon} \cdot \int f dm_0 \cdot \left(\frac{1}{(2\pi\sigma)^{d/2}}\frac{T}{(\ln T)^{d/2}}\right) \cdot Osc_+(T,\mathbf{v})$$
$$\int_0^T f(h^t(\mathbf{v}))dt \geq e^{-\varepsilon} \cdot \int f dm_0 \cdot \left(\frac{1}{(2\pi\sigma)^{d/2}}\frac{T}{(\ln T)^{d/2}}\right) \cdot Osc_-(T,\mathbf{v})$$

where $Osc_\pm(T,\mathbf{v})$ are independent of f and $Osc_+(T,\mathbf{v}) = e^{\pm\varepsilon} Osc_-(T,\mathbf{v})$ for all T large enough.

The condition $\|\xi(g^{\ln T}(\mathbf{v}))/\ln T\| < \delta$ will hold eventually for every $\mathbf{v}$ s.t. $\Xi(\mathbf{v}) = \underline{0}$. For such vectors for all T large enough

$$\frac{\int_0^T f_1(h^t(\mathbf{v}))dt}{\int_0^T f_2(h^t(\mathbf{v}))dt} = \frac{e^{\pm\varepsilon}\int f_1 dm_0 \cdot \cancel{\left(\frac{1}{(2\pi\sigma)^{d/2}}\frac{T}{(\ln T)^{d/2}}\right) \cdot \mathrm{Osc}_+(T,\mathbf{v})}}{e^{\pm\varepsilon}\int f_2 dm_0 \cdot \cancel{\left(\frac{1}{(2\pi\sigma)^{d/2}}\frac{T}{(\ln T)^{d/2}}\right) \cdot \mathrm{Osc}_-(T,\mathbf{v})}} = e^{\pm 3\varepsilon}\frac{\int f_1 dm_0}{\int f_2 dm_0}.$$

Since ε is arbitrary, $\int_0^T f_1(h^t(\mathbf{v}))dt / \int_0^T f_2(h^t(\mathbf{v}))dt \to \int f_1 dm_0 / \int f_2 dm_0$.

3.2.2 Necessity

Assume $\mathbf{v}_0$ is generic for the volume measure m_0. We show that $\Xi(\mathbf{v}_0) = \mathbf{0}$. Normalize m_0 s.t. $m_0(\widetilde{M}_0) = 1$. W.l.o.g $\xi(\mathbf{v}_0) = \mathbf{0}$.

Step 1: Egoroff Theorem. By theorem 3.2, $\frac{1}{s}\xi(g^s(\mathbf{v})) \xrightarrow[s\to\infty]{} \mathbf{0}$ m_0–a.e. By Egoroff's Theorem, the limit is uniform on a subset of almost full measure in $\widetilde{M}_0$. Thus there exists $N > 0$ s.t. $\|\frac{1}{N}\xi(g^N(\mathbf{v}))\| < \varepsilon$ on a set $\Omega_0 \subset \widetilde{M}_0$ s.t. $m_0(\widetilde{M}_0 \setminus \Omega_0) < \varepsilon$. Since $\|\frac{1}{N}\xi(g^N(\mathbf{v}))\|$ is uniformly bounded (the geodesic flow moves at finite speed),

$$\frac{1}{m_0(\widetilde{M}_0)}\int_{\widetilde{M}_0}\left(\tfrac{1}{N}\xi(g^N(\mathbf{v}))\right)dm_0(\mathbf{v})\approx\mathbf{0}.$$

Step 2: Passing to the horocycle of $\mathbf{v}_0$. Let $A_T(\mathbf{v}_0):=\{h^t(\mathbf{v}_0):0<t<T\}$, and let $\lambda_T:=$ normalized Lebesgue measure on $A_T\cap\widetilde{M}_0$, that is

$$\lambda_T:=\frac{\int_0^T 1_{\widetilde{M}_0}(h^t(\mathbf{v}_0))\delta_{h^t(\mathbf{v}_0)}dt}{\int_0^T 1_{\widetilde{M}_0}(h^t(\mathbf{v}_0))dt}\qquad(\delta=\text{Dirac's measure}).$$

Since $\mathbf{v}_0$ is generic, $\lambda_T\xrightarrow[T\to\infty]{w^*}m_0(\,\cdot\,|\widetilde{M}_0)$. So there is T_0 s.t. for all $T>T_0$,

$$\frac{1}{\lambda_T(A_T)}\int_{A_T}\left(\tfrac{1}{N}\xi(g^N(\mathbf{v}))\right)d\lambda_T(\mathbf{v})\approx\mathbf{0}.$$

Let $X_0(\mathbf{v}):=\mathbb{Z}^d$-displacement of $g^s(\mathbf{v})$ from $s=0$ to $s=N$, i.e.

$$X_0(\mathbf{v}):=\xi(g^N(\mathbf{v}))-\xi(\mathbf{v})=\xi(g^N(\mathbf{v})),$$

then what we have shown is that $\mathbb{E}_{\lambda_T}(X_0)=o(N)$.

Step 3: Epochs. Divide the *geodesic* time interval $[0,\ln T]$ into equal "epochs" of length N (up to boundary effects). Let $X_i(\mathbf{v}):=\mathbb{Z}^d$-displacement of $g^s(\mathbf{v})$ from $s=iN$ to $s=(i+1)N$, i.e.

$$X_i(\mathbf{v}):=\xi(g^{(i+1)N}(\mathbf{v}))-\xi(g^{iN}(\mathbf{v})).$$

For every $\mathbf{v}\in\widetilde{M}_0$, $\frac{1}{\ln T}\sum_{i=0}^{\lfloor\ln T/N\rfloor}X_i(\mathbf{v})=\frac{1}{\ln T}\xi(g^{\ln T}(\mathbf{v}))+O(\frac{1}{\ln T})$.

For every $\mathbf{v}\in A_T(\mathbf{v}_0)$, $\operatorname{dist}(g^{\ln T}(\mathbf{v}),g^{\ln T}(\mathbf{v}_0))\le\operatorname{diam}(g^{\ln T}(A_T(\mathbf{v}_0)))=1$, so for every $\mathbf{v}\in A_T(\mathbf{v}_0)$, $\xi(g^{\ln T}(\mathbf{v}))=\xi(g^{\ln T}(\mathbf{v}_0))+O(1)$. So

$$\frac{1}{\ln T}\sum_{i=0}^{\lfloor\ln T/N\rfloor}X_i(\mathbf{v})=\frac{1}{\ln T}\xi(g^{\ln T}(\mathbf{v}_0))+O(\tfrac{1}{\ln T}).$$

Averaging w.r.t. λ_T gives $\frac{1}{\ln T}\xi(g^{\ln T}(\mathbf{v}_0))=\frac{1}{\ln T}\sum_{i=0}^{\lfloor\ln T/N\rfloor}\mathbb{E}_{\lambda_T}(X_i)+O(\frac{1}{\ln T})$.

We already know that $\mathbb{E}_{\lambda_T}(X_0)=o(N)$. Suppose we could show that

$$\mathbb{E}_{\lambda_T}(X_i)=o(N)\text{ uniformly for }i=1,\dots,\lfloor\ln T/N\rfloor,\tag{3.2}$$

then it would follow that $\frac{1}{\ln T}\xi(g^{\ln T}(\mathbf{v}_0))=\frac{1}{\ln T}O(\frac{\ln T}{N})o(N)+O(\frac{1}{\ln T})=o_N(1)+o_T(1)$ as $T,N\to\infty$, proving that $\Xi(\mathbf{v}_0)=\mathbf{0}$. Thus the key is to show (3.2).

How to prove (3.2): Here we oversimplify a bit. The idea is to construct maps $\kappa_i:A_T(\mathbf{v}_0)\to A_T(\mathbf{v}_0)$ such that

(1) κ_i is one-to-one and measurable;

(2) κ_i is absolutely continuous with respect to Lebesgue's measure, and its Radon-Nikodym derivative is bounded away from zero and infinity, uniformly in i, N;
(3) $\kappa_i(A_T(\mathbf{v}_0)\cap[\xi=\mathbf{0}])\subset A_T(\mathbf{v}_0)\cap[\|\xi\|\leq C]$ for some C independent of i, N;
(4) κ_i "exchanges" the combinatorial behavior of the geodesic at the i-th epoch with what it does at the zeroeth epoch so that $\Xi_0\circ\kappa_i=\Xi_i+O(1)$ uniformly in i, N.

This is done using symbolic dynamics, see [SSc].

The part of $A_T(\mathbf{v}_0)$ where $\frac{1}{N}X_i\not\approx\mathbf{0}$ gets mapped by κ_i to the part of $A_T(\mathbf{v}_0)$ where $\frac{1}{N}X_0\not\approx\mathbf{0}$. This set has small measure. Since κ_i has bounded derivative, the part of $A_T(\mathbf{v}_0)$ where $\frac{1}{N}X_i\not\approx\mathbf{0}$ must also have a small measure. So $\frac{1}{N}X_i\approx\mathbf{0}$ on most of $A_T(\mathbf{v}_0)$. Averaging we get (3.2). □

3.3 Open problems

1. Extend the results of this section to regular covers of compact hyperbolic surfaces with *nilpotent* groups of deck transformations.
2. Let M be a connected orientable complete hyperbolic surface without non-constant bounded harmonic functions. In all cases I am aware of (compact surfaces, surfaces of finite area, $\mathbb{Z}^d$-covers) every $\mathbf{v}\in T^1M$ s.t.

$$\frac{1}{s}\log F(\text{base point of } g^s(\mathbf{v}))\xrightarrow[s\to\infty]{} 0 \text{ for all positive minimal eigenfunctions } F$$

has an equidistributed horocycle. How general is this phenomenon?

(The condition of absence of non-constant bounded harmonic functions is needed to guarantee the ergodicity of the horocycle flow, see §2.1.4.)

3.4 Notes and references

The references for the results in §3.1.2 are, respectively, [Fu1], [DS], [Scha], [Bu], [Ro2], and [Ra2, Ra3]. Theorem 3.2 is proposition 1.1 in [BL1]. The paper [BL1] is also the main source for the ideas and tools needed to prove Lemma 3.1. The lemma itself is proved implicitly in [LS1, Prop. 4.1] for indicators of special sets, and extends to general continuous functions with compact support by standard approximation arguments. Theorem 3.3 is proved in [SSc], and the proof we sketch here is given there in detail.

Chapter 4
Proof of the Measure Classification Theorem

We sketch the proof of the following statement: Suppose $M = \Gamma \setminus \mathbb{D}$ is a very tame surface. Then every horocycle ergodic invariant Radon measure m arises via Babillot's construction from a minimal positive eigenfunction of the Laplacian on M.[1]

4.1 It is enough to prove g–quasi-invariance

The following theorem says that it is enough to show that m is *quasi-invariant* for the geodesic flow: $m \circ g^s \sim m$ for all s, i.e. ($m(E) = 0 \Leftrightarrow m[g^s(E)] = 0$) for every s and $E \subset T^1M$ Borel measurable.

Theorem 4.1 (Babillot). *If an e.i.r.m. of the horocycle flow is geodesic quasi–invariant, then it takes the form $e^{\alpha s} d\nu ds dt$ with ν Γ–conformal with index α, and the positive eigenfunction $F(\Gamma z) = \int_{\partial\mathbb{D}} P(e^{i\theta}, z)^{\alpha} d\nu$ is minimal.*

Proof. Every measure on $T^1(\Gamma \setminus \mathbb{D})$ can be identified with a Γ–invariant measure on $T^1\mathbb{D}$. We call this measure m and describe it in the *kan*–coordinate system $T^1\mathbb{D} \cong \partial\mathbb{D} \times \mathbb{R} \times \mathbb{R}$ introduced in §1.2.1, making extensive use of Theorem 1.1.

Horocycle invariance is invariance under $h^t(e^{i\theta_0}, s_0, t_0) = (e^{i\theta_0}, s_0, t_0 + t)$. Any measure invariant under this flow has the form $d\mu(e^{i\theta}, s)dt$.

Geodesic quasi-invariance. Since $g^s \circ h^t = h^{te^{-s}} \circ g^s$, $m \circ g^s$ is h–invariant, and therefore $\frac{dm\circ g^s}{dm}$ is h–invariant. By ergodicity, $\frac{dm\circ g^s}{dm} = \text{const.}$ a.e. Call the constant $c(s)$, then it is easy to see that $c(\cdot)$ is measurable, and $c(s_1 + s_2) = c(s_1)c(s_2)$. So $c(s) = e^{\beta s}$ for some $\beta \in \mathbb{R}$ and $m \circ g^s = e^{\beta s} m \ \ (s \in \mathbb{R})$.

Since $dm = d\mu dt$ and $g^s(e^{i\theta_0}, s_0, t_0) = (e^{i\theta_0}, s_0 + s, t_0 e^{-s})$, $e^{\beta s} d\mu dt = m \circ g^s = d\mu(e^{i\theta_0}, s_0 + s)e^{-s}dt$ whence $d\mu(e^{i\theta_0}, s_0 + s) = e^{(\beta+1)s} d\mu(e^{i\theta_0}, s)$. This means that $\mu = e^{(\beta+1)s} d\nu(e^{i\theta})ds$.

Letting $\alpha := \beta - 1$ we obtain $dm = e^{\alpha s} d\nu(e^{i\theta})dsdt$ for some measure ν on $\partial\mathbb{D}$. Since m is locally finite, ν is finite.

[1] The result holds for a larger class of surfaces ("weakly tame surfaces"). See [Sa2].

Γ–invariance says that $m\circ\varphi = m$ for all $\varphi\in\Gamma$. Using the representation $(*)$ of the Γ–action in *kan*–coordinates, we obtain

$$\begin{aligned} e^{\alpha s} d\nu ds dt &= (e^{\alpha s} d\nu ds dt)\circ\varphi = e^{\alpha(s-\log|\varphi'|)}\frac{d\nu\circ\varphi}{d\nu} d\nu ds dt \\ &= \left(\frac{1}{|\varphi'|^{\alpha}}\frac{d\nu\circ\varphi}{d\nu}\right) e^{\alpha s} d\nu ds dt, \end{aligned}$$

whence $\frac{d\nu\circ\varphi}{d\nu} = |\varphi'|^{\alpha}$ for all $\varphi\in\Gamma$. So ν is Γ-conformal with parameter α.

It can be shown, using the assumption that M is tame, that the parameter α is larger than or equal to $\frac{1}{2}$ (see [Sa2], §2.3).

By Theorem 1.4, $\widetilde{F}(z) = \int_{\partial\mathbb{D}} P(e^{i\theta},z)^{\alpha} d\nu(e^{i\theta})$ is Γ-invariant, so $F(\Gamma z) := \widetilde{F}(z)$ is a well-defined positive eigenfunction on M.

F is minimal: Otherwise, it is the convex combination of non-proportional eigenfunctions. Since the Karpelevich Representation is unique when $\alpha\geq\frac{1}{2}$, ν is the convex combination of non-proportional conformal measures with exponent α. So m is the convex combination of non-proportional h–invariant Radon measures. This cannot happen for ergodic measures. ☐

Babillot's Theorem reduces the measure classification theorem to the following problem:

The problem. *Let μ be a Radon measure on $\partial\mathbb{D}\times\mathbb{R}$ which is ergodic and invariant for the* Radon-Nikodym action *of Γ*

$$\varphi(e^{i\theta},s) = (\varphi(e^{i\theta}), s-\log|\varphi'(e^{i\theta})|) \quad (\varphi\in\Gamma).$$

Let $H_\mu := \{s\in\mathbb{R} : \mu\circ g^s\sim\mu\}$. Show that $H_\mu = \mathbb{R}$ for every Γ s.t. the hyperbolic surface $\Gamma\setminus\mathbb{D}$ is very tame.

Indeed, every h-e.i.r.m. on $T^1(\Gamma\setminus\mathbb{D})$ can be identified with a Γ–invariant h–e.i.r.m. on $T^1\mathbb{D}$. By Theorem 1.1, such measures take the following form in *kan*-coordinates: $e^{\alpha s} d\mu(e^{i\theta},s)dt$, where μ is ergodic and invariant under the Radon-Nikodym action $\varphi(e^{i\theta},s) = (\varphi(e^{i\theta}), s-\log|\varphi'(e^{i\theta})|) \quad (\varphi\in\Gamma)$. Quasi-invariance under the geodesic flow is quasi-invariance under the translation flow $g^s(e^{i\theta_0},s_0) = (e^{i\theta_0}, s_0+s)$. So m is quasi invariant for the geodesic flow iff $H_\mu = \mathbb{R}$.

4.2 The possible values of H_μ and the support of e.i.r.m.

Proposition 4.1. *Let μ be an e.i.r.m for the Radon-Nikodym action on $\partial\mathbb{D}\times\mathbb{R}$, then $H_\mu = \{0\}$, $c\mathbb{Z}$, or $\mathbb{R}$.*

Proof. If $s\in H_\mu$ then $\mu\circ g^s\sim\mu$. Since both measures are invariant under the Radon-Nikodym action, $\frac{d\mu\circ g^s}{d\mu}$ is Γ–invariant. By ergodicity, $\frac{d\mu\circ g^s}{d\mu} = \text{const.}$ So

$$H_\mu = \{s \in \mathbb{R} : \mu \circ g^s = c\mu \text{ for some } c \neq 0\}.$$

It immediately follows that H_μ is a subgroup of $(\mathbb{R},+)$. We'll show that H_μ is closed, and deduce the proposition from the well known fact that the only closed subgroups of $\mathbb{R}$ are $\{0\}$, $c\mathbb{Z}$, and $\mathbb{R}$.

Suppose $s_n \in H_\mu$ and $s_n \to s_0$. Let c_n be the constants s.t. $\mu \circ g^{s_n} = c_n\mu$. Choose some $F \in C_c(\partial\mathbb{D}\times\mathbb{R})$ s.t. $\int F d\mu, \int F \circ g^{-s_0} d\mu \neq 0$. Since F is uniformly continuous, $F \circ g^{-s_n} \xrightarrow[n\to\infty]{} F \circ g^{-s_0}$ uniformly. Since F has compact support, $F \circ g^{-s_n}$ equals zero outside some fixed compact set. It follows that $c_n \int F d\mu = \int F \circ g^{-s_n} d\mu \xrightarrow[n\to\infty]{} \int F \circ g^{-s_0} d\mu$, and so $c_n \xrightarrow[n\to\infty]{} c_0 := \int F \circ g^{-s_0} d\mu / \int F d\mu \neq 0$.

The constant c_0 is independent of F (it's the limit of c_n). So for *every* $F \in C_c(\partial\mathbb{D}\times\mathbb{R})$, $\int F \circ g^{-s_0} d\mu = c_0 \int F d\mu$. It follows that $s_0 \in H_\mu$. ☐

Next, we relate H_μ to the support of μ. We begin with some abstract considerations. The RN action $\varphi \cdot (e^{i\theta}, s) = (\varphi(e^{i\theta}), s - \log|\varphi'(e^{i\theta})|)$ is a particular case of a *skew-product action*. Here is the general definition:

Skew-product actions. Suppose

- G is a countable group of measurable maps on a Borel space $(X, \mathscr{F})$
- $\Phi : G \times X \to \mathbb{R}$ is a Borel *cocycle*: $\Phi(\varphi_1\varphi_2, x) = \Phi(\varphi_2, x) + \Phi(\varphi_1, \varphi_2(x))$

The *skew product action* generated by G and Φ is the action on $X \times \mathbb{R}$ given by $\varphi(x,t) = (\varphi(x), t + \Phi(\varphi, x))$.

Radon-Nikodym Cocycle: For example, the RN action is the action generated by the Γ–action on $\partial\mathbb{D}$ and the *Radon-Nikodym cocycle*

$$R(\varphi, x) := -\log|\varphi'(x)|.$$

A measure μ on $X \times \mathbb{R}$ is called *locally finite*, if $\mu(X \times K) < \infty$ for every compact $K \subset \mathbb{R}$. Define as before $g^s(x_0, s_0) = (x_0, s_0 + s)$, and $H_\mu := \{s \in \mathbb{R} : \mu \circ g^s \sim \mu\}$.

Theorem 4.2 (Cocycle reduction theorem). *Let G be a countable group which acts on a standard space X measurably. Let $\Phi : G \times X \to \mathbb{R}$ be a measurable cocycle. If μ is a locally finite, ergodic, and invariant measure for the skew-product G–action $\varphi(x,t) = (\varphi(x), t + \Phi(\varphi, x))$ on $X \times \mathbb{R}$, then is a Borel function $u : X \to \mathbb{R}$ s.t.*

1. The set $\{(x,t) : t \in u(x) + H_\mu\}$ has full μ–measure.
2. For μ–a.e. $(x,t) \in X \times \mathbb{R}$, $\Phi(\varphi, x) + u(x) - (u \circ \varphi)(x) \in H_\mu$ for all $\varphi \in G$.
3. H_μ is contained in any closed subgroup of $\mathbb{R}$ with property 1 or with property 2.

The proof is sketched in the appendix. We remark that the cocycle reduction theorem The theorem is obvious for ergodic components of product measures on $X \times \mathbb{R}$. The point is that there is no such assumption on μ.

Applying the cocycle reduction theorem to $R(\varphi, e^{i\theta})$, we obtain the following characterization of H_μ: *It is the smallest closed subgroup of $\mathbb{R}$ s.t. μ is carried by a set of the form $\{(x,t) : t \in u(x) + H_\mu\}$.*

If H_μ were equal to $\{0\}$ or $c\mathbb{Z}$, then $\{(x,t) : t = u(x)\}$ or $\{(x,t) : t \in u(x) + c\mathbb{Z}\}$ would be left invariant mod μ by all measure preserving $\widetilde{\kappa} : \partial\mathbb{D} \times \mathbb{R} \to \partial\mathbb{D} \times \mathbb{R}$. We will construct measure preserving $\widetilde{\kappa}$ which do not leave these sets invariant, and deduce that $H_\mu = \mathbb{R}$.

The measure preserving maps we will use take the form $\widetilde{\kappa}(e^{i\theta}, s) = (\varphi_{e^{i\theta}}(e^{i\theta}), s - \log|(\varphi_{e^{i\theta}})'(e^{i\theta})|)$ where $u[\varphi_{e^{i\theta}}(e^{i\theta})] \approx u[e^{i\theta}]$ and $\log|(\varphi_{e^{i\theta}})'(e^{i\theta})|$ is "bounded away" from $c\mathbb{Z}$. To make things precise, we need the concept of *holonomies*, which is explained below.

4.3 Holonomies

Let Γ be a countable discrete subgroup of Möb($\mathbb{D}$), now viewed as a group acting on $\partial\mathbb{D}$. A *Γ–holonomy* is a map $\kappa : A \to B$ s.t.

1. $A, B \subset \partial\mathbb{D}$ are Borel
2. κ is a bi-measurable bijection
3. for all $e^{i\theta} \in A$, $\kappa(e^{i\theta}) \in \Gamma e^{i\theta}$.

A holonomy κ takes the form $\kappa(e^{i\theta}) = \varphi_{e^{i\theta}}(e^{i\theta})$, where $\varphi_{e^{i\theta}} \in \Gamma$. The map $\varphi_{e^{i\theta}}$ is unique, unless $e^{i\theta} \in \mathrm{Fix}(\Gamma) := \{e^{i\theta} : \exists \varphi \in \Gamma \setminus \{id\} \text{ s.t. } \varphi(e^{i\theta}) = e^{i\theta}\}$. This allows us to *define* for all $e^{i\theta} \notin \mathrm{Fix}(\Gamma)$

$$\kappa'(e^{i\theta}) := (\varphi_{e^{i\theta}})'(e^{i\theta})$$

Caution! $\kappa'(e^{i\theta})$ is *not* the derivative of κ, it is the derivative of the element $\varphi = \varphi_{e^{i\theta}} \in \Gamma$ s.t. $\kappa(e^{i\theta}) = \varphi(e^{i\theta})$. The holonomy itself does not need to be differentiable or even continuous.

The definition of $\kappa'(e^{i\theta})$ is ambiguous at fixed points of elements in $\Gamma \setminus \{id\}$. But this is no problem, because

Lemma 4.1. *If Γ is non–elementary and without parabolic elements (a consequence of very tameness), then $\{(e^{i\theta}, s) : \exists \varphi \in \Gamma \setminus \{id\} \text{ s.t. } \varphi(e^{i\theta}) = e^{i\theta}\}$ has zero measure for every locally finite invariant measure of the RN action of Γ.*

Proof. Assume by way of contradiction that the lemma is false. Γ is countable, so the set of fixed points of $\Gamma \setminus \{id\}$–elements is countable. Therefore, there exist $e^{i\theta_0}$, $\varphi_0 \in \Gamma \setminus \{id\}$ s.t. $\varphi_0(e^{i\theta_0}) = e^{i\theta_0}$, and $I \subset \mathbb{R}$ compact s.t. $\mu(e^{i\theta_0} \times I) > 0$.

Since φ_0 is not parabolic, $|\varphi_0'(e^{i\theta_0})| \neq 1$. This allows us to construct $M > 1$ and $n_\varphi \in \mathbb{Z}$ ($\varphi \in \Gamma \setminus \{id\}$) s.t. $|\varphi'(e^{i\theta_0})| \cdot |\varphi_0'(e^{i\theta_0})|^{n_\varphi} \in [e^{-M}, e^M]$.

Let $\Omega := \bigcup_{\varphi \in \Gamma} (\varphi \circ \varphi_0^{n_\varphi})(e^{i\theta_0} \times I)$. Ω has finite measure, since $\Omega \subset \partial\mathbb{D} \times [-M, M]$ and μ is locally finite. But $|\Gamma e^{i\theta_0}| = \infty$ ($\because \Gamma$ is non-elementary), so the union which defines Ω contains infinitely many pairwise disjoint pieces with measure $\mu(e^{i\theta_0} \times I)$. Contradiction. $\square$

Lemma 4.2. *If Γ is non–elementary and without parabolic elements (a consequence of very tameness), and if μ is invariant under the Radon-Nikodym action and $\kappa : A \to B$ is a holonomy, then $\widetilde{\kappa} : A \times \mathbb{R} \to B \times \mathbb{R}$ defined by $(x,s) \mapsto (\kappa(x), s - \log|\kappa'(x)|)$ preserves μ.*

Proof. Let $A_\varphi := \{e^{i\theta} : \kappa(e^{i\theta}) = \varphi(e^{i\theta}), e^{i\theta} \not\in \mathrm{Fix}(\Gamma)\}$ $(\varphi \in \Gamma)$.

For every $E \subset A \times \mathbb{R}$, $\widetilde{\kappa}(E) = \bigcup_{\varphi\in\Gamma} \widetilde{\kappa}(E \cap (A_\varphi \times \mathbb{R}))$. The union is disjoint, because $\widetilde{\kappa}(E \cap (A_\varphi \times \mathbb{R})) \subset \kappa(A_\varphi) \times \mathbb{R}$, A_φ are pairwise disjoint, and κ is one-to-one. Since $\mu[\widetilde{\kappa}(E \cap (A_\varphi \times \mathbb{R}))] = (\mu \circ \varphi)[E \cap (A_\varphi \times \mathbb{R})] = \mu[E \cap (A_\varphi \times \mathbb{R})]$, we get $\mu[\widetilde{\kappa}(E \cap (A_\varphi \times \mathbb{R}))] = \mu[E \cap (A_\varphi \times \mathbb{R})]$. Summing over $\varphi \in \Gamma$ gives $\widetilde{\kappa}(E) = E$. □

4.3.1 Using holonomies to show that $H_\mu \neq \{0\}$

Suppose $\Gamma \setminus \mathbb{D}$ is very tame, and μ is a locally finite measure which is ergodic and invariant for the Radon Nikodym action of Γ on $\partial\mathbb{D}$.

Lemma 4.3 (First Holonomy Lemma). *There are constants $M, S > 0$ s.t. for every $\varepsilon > 0, n \in \mathbb{N}$ there is a holonomy $\kappa : A \to B$ with the following properties:*

1. $A, B \subset \partial\mathbb{D}$ are measurable and $A \times \mathbb{R}$ has full μ measure;
2. $|\kappa(e^{i\theta}) - e^{i\theta}| < \varepsilon$ for all $e^{i\theta} \in A$;
3. $-\log|\kappa'(e^{i\theta})| \in nS + [-M, M]$ for all $e^{i\theta} \in A$.

We will give the proof later. First we'll show how to use the lemma to show that $H_\mu \neq \{0\}$. Assume by way of contradiction that $H_\mu = \{0\}$.

Step 1: μ is non-atomic. Otherwise, $\mu\{(e^{i\theta_0}, s_0)\} > 0$ for some θ_0, s_0. We know already that $e^{i\theta_0}$ is not a fixed point of some $\varphi \in \Gamma \setminus \{id\}$.

Use the first holonomy lemma with $\varepsilon_j \to 0$ and $n > M/S$ to construct holonomies κ_j s.t. $|\kappa_j(e^{i\theta_0}) - e^{i\theta_0}| \xrightarrow[j\to\infty]{} 0$ and $nS - M \leq -\log|\kappa_j'(e^{i\theta_0})| \leq nS + M$.

Since $-\log|\kappa_j'(e^{i\theta_0})| \neq 0$, $\kappa_j(e^{i\theta_0}) = \varphi(e^{i\theta_0})$ for $id \neq \varphi \in \Gamma$. Since $e^{i\theta_0}$ is not a fixed point, $\kappa_j(e^{i\theta_0}) \neq e^{i\theta_0}$.

Since $0 < |\kappa_j(e^{i\theta_0}) - e^{i\theta_0}| \xrightarrow[j\to\infty]{} 0$, $\{\kappa_j(e^{i\theta_0}) : j \in \mathbb{N}\}$ is an infinite set. So $E = \{(\kappa_j(e^{i\theta_0}), s_0 - \log|\kappa_j'(e^{i\theta_0})|) : j \in \mathbb{N}\}$ is infinite. By invariance, $\mu(E) = \infty$. But $E \subset \partial\mathbb{D} \times (s_0 + nS + [-M, M])$ and μ is locally finite, so $\mu(E) < \infty$.

Step 2: μ is carried by the graph of a bounded function. By Theorem A.2, and since $H_\mu = \{0\}$, there is a Borel function $u : \partial\mathbb{D} \to \mathbb{R}$ s.t. $\mathrm{graph}(u) := \{(e^{i\theta}, u(e^{i\theta})) : e^{i\theta} \in \partial\mathbb{D}\}$ has full measure. We have to show that $\operatorname{ess\,sup}|u| < \infty$.

Assume by way of contradiction that $\operatorname{ess\,sup}|u| = \infty$. Let S, M be as in the first holonomy lemma. Since $\operatorname{ess\,sup}|u| = \infty$, there is an interval I of length S s.t. $\mu(\partial\mathbb{D} \times I) \neq 0$ and $\mu(\partial\mathbb{D} \times (n + I)) \neq 0$ for some $|n| \gg M/S$. Let $J := N_M(nS + I)$ (the M–neighborhood of $nS + I$), then $I \cap J = \varnothing$ (figure 4.1)

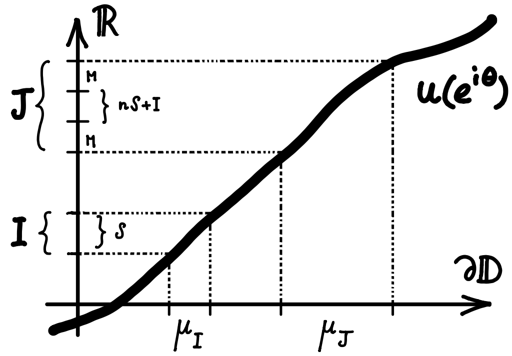

Fig. 4.1 Construction of I, J, μ_I, μ_J

Define two measures μ_I, μ_J on $\partial\mathbb{D}$ by $\mu_I(E) := \mu(E \times I)$ and $\mu_J(E) := \mu(E \times J)$. These are finite positive measures, and $\mu_I \perp \mu_J$, because $u \in I$ μ_I–a.e., and $u \in J$ μ_J–almost everywhere.

Since μ_I, μ_J are singular, there exists a closed arc A s.t. $0 \neq \mu_J(A) < \frac{1}{2}\mu_I(A)$. [If this fails for all closed arcs, $\mu_I \ll \mu_J$.] Since A is closed, for every ε small enough, the ε-neighborhood $N_\varepsilon(A)$ satisfies $0 \neq \mu_J[N_\varepsilon(A)] < \frac{2}{3}\mu_I(A)$.

The holonomy lemma provides a Γ–holonomy κ s.t. $|\kappa(e^{i\theta}) - e^{i\theta}| < \varepsilon$ and $-\log|\kappa'| \in nS + [-M, M]$. Let $\widetilde{\kappa}(e^{i\theta}, s) = (\kappa(e^{i\theta}), s - \log|\kappa'(e^{i\theta})|)$, then

$$\widetilde{\kappa}(A \times I) \subset N_\varepsilon(A) \times N_M(nS + I) \subset N_\varepsilon(A) \times J.$$

Since $\widetilde{\kappa}$ is measure preserving, $\mu_I(A) = \mu(A \times I) = \mu[\widetilde{\kappa}(A \times I)] \leq \mu_J[N_\varepsilon(A)] < \frac{2}{3}\mu_I(A)$. Dividing by $\mu_I(A)$, we get $1 < \frac{2}{3}$, a contradiction. So $\operatorname{ess\,sup}|u| < \infty$.

Step 3: completion of the proof. The last step shows that μ is carried by the graph of a bounded function. So μ is carried by a bounded set, say $\partial\mathbb{D} \times [-B, B]$.

Choose n so large that $nS > 2B + M + 1$, and let κ be the Γ–holonomy s.t. $-\log|\kappa'| \in nS + [-M, M]$. Then $\widetilde{\kappa}(\partial\mathbb{D} \times [-B, B]) \subset (\partial\mathbb{D} \times [-B, B])^c$, so $\widetilde{\kappa}$ maps a set of full measure into a set of zero measure. But $\widetilde{\kappa}$ is measure preserving, so we obtain a contradiction. This shows that it is impossible for H_μ to equal $\{0\}$. □

4.3.2 Using holonomies to show that $H_\mu \neq c\mathbb{Z}$

Lemma 4.4 (Second Holonomy Lemma). *Fix* $c \neq 0$. *There are constants* $M > 0$, $0 < \delta < |c|$ *s.t. for every* $\varepsilon > 0$ *there is a holonomy* $\kappa : A \to B$ *s.t.:*

1. $A, B \subset \partial\mathbb{D}$ *are measurable and* $A \times \mathbb{R}$ *has full* μ *measure;*
2. $|\kappa(e^{i\theta}) - e^{i\theta}| < \varepsilon$ *for all* $e^{i\theta} \in A$*;*
3. $-\log|\kappa'(e^{i\theta})| \in [-M, M]$ *for all* $e^{i\theta} \in A$*;*
4. $-\log|\kappa'(e^{i\theta})|$ *is within distance at least* δ *away from* $c\mathbb{Z}$ *for all* $e^{i\theta} \in A$.

We will give the proof later. First we'll show how use this to show that $H_\mu \neq c\mathbb{Z}$. Assume by way of contradiction that $H_\mu = c\mathbb{Z}$ for $c \neq 0$.

Step 1: The structure of μ. We claim that $\exists u : \partial\mathbb{D} \to \mathbb{R}$ bounded Borel s.t. the change of coordinates $\vartheta(x,s) = (x, s - u(x))$ transforms μ into the form

$$\mu \circ \vartheta^{-1} = e^{\alpha s} d\nu(e^{i\theta}) dm_{c\mathbb{Z}}$$

where $\alpha \in \mathbb{R}$, $m_{c\mathbb{Z}}$ =counting measure, and ν is a finite measure on $\partial\mathbb{D}$ s.t. for some constant C, $\frac{1}{C}|\varphi'|^\alpha \leq \frac{d\nu\circ\varphi}{d\nu} \leq C|\varphi'|^\alpha$ for all $\varphi \in \Gamma$.

Proof. By Theorem A.2, if $H_\mu = c\mathbb{Z}$, then μ is carried by $\{(e^{i\theta}, t) : t \in u(e^{i\theta}) + c\mathbb{Z}\}$ for some Borel measurable $u : \partial\mathbb{D} \to \mathbb{R}$. Passing to $u(\text{mod } c\mathbb{Z})$, we may take u to be bounded.

Define $\vartheta : \partial\mathbb{D} \times \mathbb{R} \to \partial\mathbb{D} \times \mathbb{R}$ by $\vartheta(e^{i\theta}, t) = (e^{i\theta}, t - u(e^{i\theta}))$, then $\mu \circ \vartheta^{-1}$ is carried by $\partial\mathbb{D} \times c\mathbb{Z}$.

Since $H_\mu = \{s : \exists c(s) \text{ s.t. } \mu \circ g^s = c(s)\mu\} = c\mathbb{Z}$, $\mu \circ g^c = e^\alpha \mu$ for some $\alpha \in \mathbb{R}$. It follows that $(\mu \circ \vartheta^{-1}) \circ g^c = e^\alpha (\mu \circ \vartheta^{-1})$, so $e^{-\alpha s} d\mu \circ \vartheta^{-1}(e^{i\theta}, s)$ is invariant under the action of $c\mathbb{Z}$ by translation on the second coordinate. This forces $\mu \circ \vartheta^{-1} = e^{\alpha s} d\nu(e^{i\theta}) dm_{c\mathbb{Z}}(s)$ for some measure ν on $\partial\mathbb{D}$.

Since u is bounded, ϑ preserves local finiteness. So ν is finite. Playing with the Γ–invariance of μ, we find that $\frac{d\nu\circ\varphi}{d\nu} = |\varphi'|^\alpha e^{-\alpha(u - u\circ\varphi)}$ (exercise, see the proof of Babillot's theorem). Since $|u|$ is bounded, $\frac{d\nu\circ\varphi}{d\nu} \asymp |\varphi'|^\alpha$. □

Step 2: Applying the holonomy lemma. We construct a set $E \subset \partial\mathbb{D}$ of positive ν measure, $\delta > 0$, and a Γ–holonomy κ s.t. on E, $-\log|\kappa'| + u - u\circ\kappa$ is at least $\delta/2$ units of distance away from $c\mathbb{Z}$.

Find $u_0 \in \mathbb{R}$ s.t. $U := [|u - u_0| < \frac{1}{4}\delta]$ has positive ν measure. Fix a small constant θ and choose a closed arc A s.t. $\nu[A \cap U] > (1-\theta)\nu(A)$.

Since A is closed, there exists $\varepsilon > 0$ so small that $\nu[A \cap U] > (1 - 2\theta)\nu[N_\varepsilon(A)]$. By the second holonomy lemma, there exists a Γ–holonomy κ s.t.

1. $|\kappa(e^{i\theta}) - e^{i\theta}| < \varepsilon$
2. $-\log|\kappa'| \in [-M, M]$
3. $\text{dist}(-\log|\kappa'|, c\mathbb{Z}) > \delta$

where $M, \delta > 0$ are independent of ε. Let $E := (A \cap U) \cap \kappa^{-1}(A \cap U)$.

We show how to choose θ to guarantee that $\nu(E) > 0$. For every $\varphi \in \Gamma$, $\frac{d\nu\circ\varphi}{d\nu} \geq C^{-1}|\varphi'|^\alpha$ for all $\varphi \in \Gamma$ so

$$\nu(A\cap U)+(\nu\circ\kappa)(A\cap U)=\int_{A\cap U}\left(1+\frac{d\nu\circ\kappa}{d\nu}\right)d\nu\geq\nu[A\cap U]\left(1+C^{-1}e^{-M\alpha}\right)$$
$$>(1-2\theta)\left(1+C^{-1}e^{-M\alpha}\right)\nu[N_\varepsilon(A)].$$

For θ small, this is more than $\nu[N_\varepsilon(A)]$. Since by construction $A\cap U,\kappa(A\cap U)\subset N_\varepsilon(A)$, $\nu[(A\cap U)\cap\kappa(A\cap U)]>0$. Since ν is non-singular w.r.t. Γ, $\nu[E]>0$.

On E, $|u-u_0|<\frac{1}{4}\delta$ and $|u_0-u\circ\kappa|<\frac{1}{4}\delta$, so $|u-u\circ\kappa|<\frac{1}{2}\delta$. Since by construction, $\mathrm{dist}\,(-\log|\kappa'|,c\mathbb{Z})>\delta$, we have $\mathrm{dist}\,(-\log|\kappa'|+u-u\circ\kappa,c\mathbb{Z})>\frac{1}{2}\delta$. □

Step 3: $H_\mu\neq c\mathbb{Z}$**.** Let $\Omega:=\vartheta^{-1}(E\times c\mathbb{Z})$ and $\widetilde{\kappa}(x,s)=(\kappa(x),s-\log|\kappa'(x)|)$, where E and κ were constructed above, then

$$\begin{aligned}0&\neq\mu(\Omega)=(\mu\circ\widetilde{\kappa})(\Omega)=(\mu\circ\vartheta^{-1})(\vartheta\circ\widetilde{\kappa}\circ\vartheta^{-1}(E\times c\mathbb{Z}))\\&\leq(\mu\circ\vartheta^{-1})\{(\kappa(e^{i\theta}),\underbrace{s-\log|\kappa'(e^{i\theta})|+u(e^{i\theta})-(u\circ\kappa)(e^{i\theta})}_{\frac{1}{2}\delta\text{–away from }c\mathbb{Z}}):e^{i\theta}\in E,s\in c\mathbb{Z}\}\\&\leq(\mu\circ\vartheta^{-1})[\partial\mathbb{D}\times(\mathbb{R}\setminus c\mathbb{Z})]=0\text{ because }\mu\circ\vartheta^{-1}=e^{\alpha s}d\nu dm_{c\mathbb{Z}}.\end{aligned}$$

This contradiction shows that it is impossible that $H_\mu=c\mathbb{Z}$. □

4.4 The equation $R(\cdot,e^{i\theta})\approx t$

We have reduced the proof the main result to the proof of the holonomy lemmas. To prove these lemmas, we need to be able to solve systems of the form

$$R(\varphi,e^{i\theta}):=-\log|\varphi'(e^{i\theta})|\approx t\text{ and }\varphi(e^{i\theta})\approx e^{i\theta}\qquad(\ddagger)$$

Here $e^{i\theta}\in\partial\mathbb{D},t\in\mathbb{R}$ are given, and $\varphi\in\Gamma$ is the unknown that we need to find.

This is where we need to wrestle with the infinite genus of $\Gamma\setminus\mathbb{D}$: If $\Gamma\setminus\mathbb{D}$ is "large," then Γ is "small," and there is no reason why

$$\bigcup_{\varphi\in\Gamma}\{e^{i\theta}:R(\varphi,e^{i\theta})\in[t-\varepsilon,t+\varepsilon]\}$$

should cover $\partial\mathbb{D}$. If it does not, (‡) may have no solution.

The trick is to show that e.i.r.m. are carried by $(e^{i\theta},*)$ such that system (‡) can be solved for a bounded set of t's which generates a dense subgroup of $\mathbb{R}$. To do this we will approximate the Radon Nikodym cocycle by another cocycle, which is easier to tie to the geometric assumptions on $\Gamma\setminus\mathbb{D}$.

4.4.1 The Busemann cocycle $B(\gamma, \widetilde{\gamma})$

Suppose $M = \Gamma \setminus \mathbb{D}$, let $g : T^1M \to T^1M$ be the *geodesic* flow, and let $\mathfrak{B}$ denote the space of oriented geodesics on M, viewed as subsets of T^1M.

The homoclinic equivalence relation: This is the equivalence relation on $\mathfrak{B}$ given by $\gamma_1 \sim \gamma_2$ iff $\exists \omega_i, \omega_i^* \in \gamma_i$ $(i = 1, 2)$ s.t.

1. $\operatorname{dist}(g^{-s}\omega_1, g^{-s}\omega_2) \xrightarrow[s\to\infty]{} 0$, and
2. $\operatorname{dist}(g^{s}\omega_1^*, g^{s}\omega_2^*) \xrightarrow[s\to\infty]{} 0$.

Busemann's cocycle: The function $B : \{(\gamma_1, \gamma_2) \in \mathfrak{B} \times \mathfrak{B} : \gamma_1 \sim \gamma_2\} \to \mathbb{R}$ given informally by $B(\gamma_1, \gamma_2) =$ "length(γ_2) − length(γ_1)."

Formally, $B(\gamma_1, \gamma_2) := \operatorname{dist}_{\gamma_2}(\omega_2, \omega_2^*) - \operatorname{dist}_{\gamma_1}(\omega_1, \omega_1^*)$ whenever ω_i, ω_i^* are as above (Fig. 4.2). Here and throughout, $\operatorname{dist}_\gamma(\omega, \omega') :=$unique s s.t. $g^s(\omega) = \omega'$. The definition is independent of the choice of ω_i, ω_i^* (exercise).

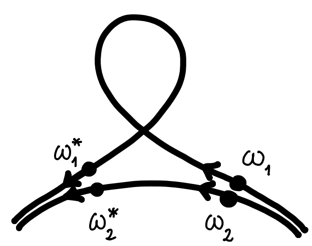

Fig. 4.2 Busemann's cocycle

Let $\gamma[e^{i\theta_1}, e^{i\theta_2}]$ denote the geodesic in $\mathbb{D}$ which starts at $e^{i\theta_1}$ and ends at $e^{i\theta_2}$. It projects to the geodesic $\Gamma\gamma[e^{i\theta_1}, e^{i\theta_2}]$ on $\Gamma \setminus \mathbb{D}$.

Two geodesics on $\mathbb{D}$ are backward (resp. forward) asymptotic iff they have the same beginning (resp. end) point. Two geodesics on $\Gamma \setminus \mathbb{D}$ are backward (resp. forward) asymptotic if they have lifts to $\mathbb{D}$ which are backward (resp. forward) asymptotic. Consequently, $\big(\Gamma\gamma[e^{i\alpha_1}, e^{i\alpha_2}] \sim \Gamma\gamma[e^{i\beta_1}, e^{i\beta_2}]\big) \Leftrightarrow \big(e^{i\alpha_i} \in \Gamma e^{i\beta_i}\ (i = 1, 2)\big)$.

Theorem 4.3 (Approximation Property). *If* $\varphi \in \Gamma$ *and* $|\varphi(e^{i\theta}) - e^{i\theta}| < 1$*, then*

$$|R(\varphi, e^{i\theta}) - B(\gamma, \widetilde{\gamma})| \le 4|\varphi(e^{i\theta}) - e^{i\theta}|^2 \text{ where } \begin{cases} \gamma := \Gamma\gamma[-e^{i\theta}, e^{i\theta}] \\ \widetilde{\gamma} := \Gamma\gamma[-e^{i\theta}, \varphi(e^{i\theta})]. \end{cases}$$

The proof is given in the appendix.

4.4.2 Cutting sequences and cut'n'paste constructions

We develop tools for solving $B(\gamma,\cdot)\approx t$ for geodesics γ which cross a pair of pants (pop) Y. Since the Busemann cocycle approximates the Radon-Nikodym cocycle, this will help us to solve $R(\cdot,e^{i\theta})\approx t$ (the main ingredient in the proof of the holonomy lemmas).

Hyperbolic octagons. Suppose $Y\subset\Gamma\setminus\mathbb{D}$ is a pair of pants ("pop"), see §1.3.1. The *seams* of Y are the shortest geodesic segments connecting its boundary components. Choose two seams out of the three. Label them on both sides, one by $\alpha,\overline{\alpha}$ and the other by $\beta,\overline{\beta}$. If we cut a Y along these seams and lift to $\mathbb{D}$, we obtain a hyperbolic octagon O_Y all whose angles equal 90° (Fig. 1.3).

The two seams of Y we cut lift to four geodesic arcs labeled by $\alpha,\overline{\alpha}$ and $\beta,\overline{\beta}$, and the boundary components lift to four geodesic arcs a,b_1,b_2,c which intersect them at right angles. (One boundary component lifts to $b_1\cup b_2$.) Put the labels *outside* O_Y, and call them *external labels*. The external labels of the same seam are inverse to each other, and they represent the two sides of the seam on Y.

Since $\alpha=\overline{\alpha}\mod\Gamma,\beta=\overline{\beta}\mod\Gamma$, there are (unique) $\varphi_\alpha,\varphi_\beta\in\Gamma$ s.t. $\varphi_\alpha[\alpha]=\overline{\alpha}$ and $\varphi_\beta[\beta]=\overline{\beta}$. Let $\varphi_{\overline{\alpha}}:=\varphi_\alpha^{-1}$, $\varphi_{\overline{\beta}}:=\varphi_\beta^{-1}$.

Extend the labeling scheme to $\bigcup_{\varphi\in\langle\varphi_\alpha,\varphi_\beta\rangle}\varphi(O_Y)$. Now every lifted seam has two labels, one internal and one external.

Cutting sequences (Artin). Suppose a vector ω has base point in Y. The *cutting sequence* of ω is the ordered list of labels $x_i\in\{\alpha,\overline{\alpha},\beta,\overline{\beta}\}$ of the seams crossed by the geodesic ray of ω, subject to the convention that crossing from side x to side $\bar{x}$ is denoted by x and crossing from side $\bar{x}$ to side x is denoted by $\bar{x}$.

The cutting sequence may be empty, finite, or infinite. It is always *reduced*: $(x_i,x_{i+1})\neq(\alpha,\overline{\alpha}),(\beta,\overline{\beta}),(\overline{\beta},\beta),(\overline{\alpha},\alpha)$. The converse also holds: every finite reduced word can be realized as part of a cutting sequence of some $\omega\in T^1Y$ [Se1].

If ω has cutting sequence $x_1x_2\cdots x_N$ and we lift ω to a vector in O_Y, then the lift of the geodesic of ω

- leaves O_Y through the side with internal label x_1, and enters $\varphi_{x_1}(O_Y)$,
- leaves $\varphi_{x_1}(O_Y)$ through the side with internal label x_2, and enters $(\varphi_{x_1}\circ\varphi_{x_2})(O_Y)$,

....

- leaves $(\varphi_{x_1}\circ\cdots\circ\varphi_{x_{N-1}})(O_Y)$ through the side with internal label x_N and enters $(\varphi_{x_1}\circ\cdots\circ\varphi_{x_N})(O_Y)$.

See Fig. 4.3. The converse is also true: if γ is the projection of a geodesic emanating from a point p inside O_Y and terminating at a point $e^{i\theta}$ which lies "under" an edge with internal label x, and $(x_1,\ldots,x_N,x)$ is reduced, then the geodesic connecting p to $(\varphi_{x_1}\circ\cdots\circ\varphi_{x_N})(e^{i\theta})$ has cutting sequence starting with $x_1x_2\cdots x_N$.

Cut'n'paste. Suppose γ crosses a pop Y with cutting sequence $\underline{u}\,\underline{v}$, and suppose $\underline{w}$ is a word s.t. the concatenation $\underline{u}\,\underline{w}\,\underline{v}$ is reduced. We describe a procedure for generating a geodesic $\gamma'\sim\gamma$ whose cutting sequence is up to edge effects $\underline{u}\,\underline{w}\,\underline{v}$:

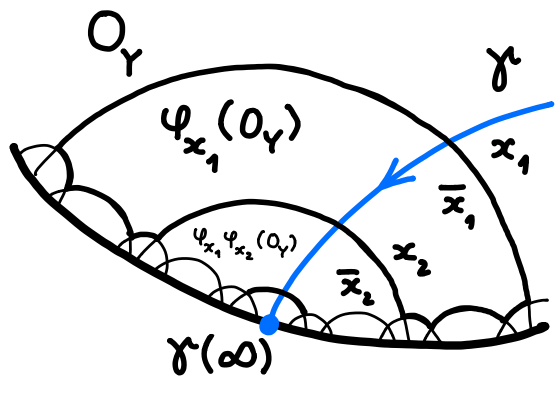

Fig. 4.3 A geodesic with cutting sequence $(x_1, x_2, x_3, \ldots)$

1. Mark a point p on γ s.t. the cutting sequence up to p is $\underline{u}$ and the cutting sequence from p is $\underline{v}$
2. Lift γ to $\mathbb{D}$ so that p lifts to a point in O_Y, the hyperbolic octagon of Y. Call the lift $\widetilde{\gamma}$, and denote its endpoints by $\gamma(\pm\infty) \in \partial\mathbb{D}$.
3. Let $\widetilde{\gamma}^*$ denote the geodesic s.t. $\widetilde{\gamma}^*(-\infty) = \gamma(-\infty)$ and $\widetilde{\gamma}^*(\infty) = \varphi_{\underline{w}}[\widetilde{\gamma}(\infty)]$, where $\varphi_{\underline{w}} = \varphi_{x_1} \circ \cdots \circ \varphi_{x_N}$ and $\underline{w} = x_1 x_2 \cdots x_N$.
4. Project $\widetilde{\gamma}^*$ to a geodesic γ^* on $\Gamma \setminus \mathbb{D}$.

The geodesic γ^* is homoclinic to γ, and has cutting sequence $\underline{u}\,\underline{w}\,\underline{v}$.

Notice the by-product $\varphi_{\underline{w}}$ of this construction. This is an element of Γ. We call $\varphi_{\underline{w}}$ the *"isometry which pastes* $\underline{w}$" (to γ). "Cutting" can be achieved in a similar way, by applying the inverse of $\varphi_{\underline{w}}$ (to $\varphi_{\underline{w}}(\gamma)$).

Remark: The possibility to generate one orbit from another by means of a cut'n'paste construction is a reflection of the local product structure of the geodesic flow on a hyperbolic surface. Similar constructions are also possible for other Anosov flows, such as geodesic flows on compact surfaces with variable negative curvature.

4.4.3 The derivative of cut'n'paste isometries

A word $\underline{w}$ will be called *prime*, if $\underline{w} \neq \underline{u}^k$ for $k > 1$ and $\underline{w}\,\underline{w}$ is reduced. Then $(\cdots \underline{w}, \underline{w}, \underline{w}, \cdots)$ is the cutting sequence of a periodic geodesic in Y. We denote the length of this closed geodesic $p_Y(\underline{w})$.

Lemma 4.5 (Basic estimate). *Suppose γ crosses a pop Y with cutting sequence $\underline{u}\,\underline{w}^{2k}\underline{v}$ with $\underline{w}$ prime. Let γ' denote the geodesic obtained by "pasting" another n $\underline{w}$'s: $\underline{u}\,\underline{w}^{2k+n}\,\underline{v}$. Then $B(\gamma,\gamma') \approx np_Y(\underline{w})$. More precisely,*

$$|B(\gamma,\gamma') - np_Y(\underline{w})| < 100e^{\mathrm{diam}(O_Y)}e^{-(k-2)p_Y(\underline{w})}.$$

Proof (general idea, details in [Sa2]). Let p be the point on γ such that the cutting sequence of γ up to p is $\underline{u}\,\underline{w}^k$ and the cutting sequence after p is $\underline{w}^k\underline{v}$.

Let $\omega_1 = \omega_1^*$ denote the velocity vector of γ at p. Let ω_2, ω_2^* denote the velocity vectors of γ' s.t. $d(g^{-s}\omega_1, g^{-s}\omega_2) \xrightarrow[s\to\infty]{} 0$ and $d(g^{s}\omega_1^*, g^{-s}\omega_2^*) \xrightarrow[s\to\infty]{} 0$, then $B(\gamma,\gamma') = \mathrm{dist}_{\gamma'}(\omega_2, \omega_2^*)$.

Consider the geodesic arc $A := (\omega_2, \omega_2^*)$. The proof goes by showing that A is very close to n–windings of the closed geodesic with cutting sequence $\underline{w}^\infty$. ☐

The lemma provides a sufficient condition on $e^{i\theta}$ and t for the existence of a solution $\varphi \in \Gamma$ to the equation $R(\varphi, e^{i\theta}) \approx t$: Let $\omega(e^{i\theta})$:=unit vector based in $0 \in \mathbb{D}$ and pointing at $e^{i\theta} \in \partial\mathbb{D}$, and let $\Gamma\omega(e^{i\theta})$ be its projection to $\Gamma \setminus \mathbb{D}$.

The condition: *$t = np_Y(\underline{w})$ and the geodesic ray of $\Gamma\omega(e^{i\theta})$ crosses a pop Y where its cutting sequence has the form $\underline{u}\,\underline{w}^{2k}\underline{v}$ with $k \to \infty$.*

If this happens, and φ is the cut'n'paste isometry which replaces $\underline{w}^{2k}$ by $\underline{w}^{2k+n}$, then $\varphi \in \Gamma$, $|\varphi(e^{i\theta}) - e^{i\theta}| = o(1)$ and $R(\varphi, e^{i\theta}) = np_Y(\underline{w}) + o(1)$, as $k \to \infty$. The little ohs can be shown to be uniform over all pop whose boundary components all have lengths in $[L^{-1}, L]$.

We shall see below that this condition holds for a.e. geodesic with respect to any ergodic invariant Radon measure on a very tame surface. This is the key to the proof of the first holonomy lemma (details below). The second holonomy lemma requires more: We need to know that the collection of t generates a "sufficiently dense" subset of $\mathbb{R}$. This information is the content of the following lemma.

Lemma 4.6 (Aperiodicity). *Fix L. For every $c > 0$ there exists a finite collection of reduced prime words $\mathscr{W}(c,L)$ and $\delta = \delta(c) > 0$ s.t. for every pop Y whose boundary lengths all belong to $[L^{-1}, L]$ there exists $\underline{w} \in \mathscr{W}(c,L)$ s.t. $\mathrm{dist}(p_Y(\underline{w}), c\mathbb{Z}) > \delta$.*

Proof (sketch, details in [Sa2]). For a single pop Y, the existence of such a finite set of words is a consquence of the aperiodicity of the length spectrum of hyperbolic surfaces (Dalb'o [Dal], Guivarc'h & Raugi [GR]).

The number $p_Y(\underline{w})$ is a continuous function of the boundary lengths of Y. Since we are assuming that these lengths belong to $[L^{-1}, L]$, the proof can be completed using a compactness argument. ☐

4.5 Proof of the holonomy lemmas

4.5.1 *Scenarios when* $R(\cdot, e^{i\theta}) \approx t$ *can be solved for many* t

Suppose M is very tame, then M has a pants decomposition $\{Y_i\}$ s.t. the lengths of all the boundary components belong to $[L^{-1}, L]$.

Fix $c > 0$ and let $\mathscr{W}(c,L)$ be a finite set of prime reduced words $\underline{w}$ as in Lemma 4.6. Then $\exists \delta, M > 0$ s.t. for every pop Y with boundary component lengths in $[L^{-1}, L]$ there exists at least one $\underline{w} \in \mathscr{W}(c,L)$ s.t. $\mathrm{dist}(p_Y(\underline{w}), c\mathbb{Z}) > \delta$, and at least one $\underline{w}$ s.t. $p_Y(\underline{w}) \in [M^{-1}, M]$.

Write $\mathscr{W}(c,L) = \{\underline{w}_1, \ldots, \underline{w}_N\}$ and form, for every k, the concatentation

$$\underline{w}^{(2k)} := \underline{w}_1^{2k}\, u_1\, \underline{w}_2^{2k}\, u_2\, \underline{w}_3^{2k} \cdots u_{N-1}\, \underline{w}_N^{2k} \tag{4.1}$$

where $u_i \in \{\alpha, \overline{\alpha}, \beta, \overline{\beta}\}$ are chosen once and for all to make $\underline{w}^{(2k)}$ reduced. We will consider the following two scenarios:

Scenario I: *For every $k_0 \in \mathbb{N}$ and $\underline{w} \in \mathscr{W}(c,L)$, the geodesic ray $\Gamma\omega(e^{i\theta})$ enters a pop Y_i in the pants decomposition of M with cutting sequence $\underline{u}\,\underline{w}^{(2k)}\,\underline{v}$ with $k > k_0$ and $\underline{v}$* **finite**. *(The ray eventually leaves Y.)*

Scenario II: *For every $k_0 \in \mathbb{N}$ and $\underline{w} \in \mathscr{W}(c,L)$, the geodesic ray $\Gamma\omega(e^{i\theta})$ enters a pop Y_i in the pants decomposition of M with cutting sequence $\underline{u}\,\underline{w}^{(2k)}\,\underline{v}$ with $k > k_0$ and $\underline{v}$* **infinite**. *(The ray gets trapped in Y for ever.)*

In these scenarios we always have room to "paste" the $\underline{w} \in \mathscr{W}(c,L)$ which serves our purposes. The isometry $\varphi \in \Gamma$ which implements the pasting satisfies $R(\varphi, e^{i\theta}) \in [M^{-1}, M]$ or $\mathrm{dist}(R(\varphi, e^{i\theta}), c\mathbb{Z}) > \delta$, which is what we need to prove the holonomy lemmas.

4.5.2 *One of the scenarios happens almost surely*

Let M be a very tame surface and let $\mathscr{Y}$ be the collection of pops in the pants decomposition of M. Since M is very tame, there is a constant $L > 1$ s.t. for every $Y \in \mathscr{Y}$, all the boundary components of Y have lengths in $[L^{-1}, L]$.

Either a geodesic ray enters some $Y \in \mathscr{Y}$ where it has an infinite cutting sequence, or it enters and leaves ("crosses") infinitely many $Y_i \in \mathscr{Y}$ through different boundary components (we do not claim that the Y_i are distinct). Therefore the following properties of $(e^{i\theta}, s)$ are mutually exclusive:

(A) The geodesic ray of $\Gamma g^s[\omega(e^{i\theta})]$ enters and leaves pops in $\mathscr{Y}$ through different boundary components, infinitely many times
(B) The geodesic ray of $\Gamma g^s[\omega(e^{i\theta})]$ is eventually trapped in the union of two adjacent pops.

The sets $\{(e^{i\theta},s)\in\partial\mathbb{D}\times\mathbb{R}:\text{(A) happens}\}$ and $\{(e^{i\theta},s)\in\partial\mathbb{D}\times\mathbb{R}:\text{(B) happens}\}$ are invariant under the Radon-Nikodym action of Γ, because by Theorem 1.1, the geodesic ray of $\Gamma\omega(e^{i\theta})$ and the geodesic ray of $\Gamma g^{s-\log|\varphi'(e^{i\theta})|}\omega(\varphi(e^{i\theta}))$ sit on the same horocycle, and are therefore forward asymptotic. So for every ergodic invariant Radon measure μ of the Radon-Nikodym action on $\partial\mathbb{D}\times\mathbb{R}$, either property (A) holds μ–a.e., or (B) holds μ–a.e.

If (A) holds a.e. then scenario I holds a.e.: Let μ be an ergodic invariant Radon measure for the Radon-Nikodym action on $\partial\mathbb{D}\times\mathbb{R}$. Assume by way of contradiction that scenario I fails for a set of positive measure of $(e^{i\theta},s)$, then there are $k_0\in\mathbb{N}$ and $I\subset\mathbb{R}$ compact s.t. the following set has positive measure:

$$\Omega := \left\{ (e^{i\theta},s)\in\partial\mathbb{D}\times I \;\middle|\; \begin{array}{l}\text{The geodesic ray of } \Gamma g^{s}\omega(e^{i\theta}) \text{ does not intersect}\\ Y\in\mathscr{Y} \text{ where its cutting sequence contains } \underline{w}^{(2k_0)}\end{array} \right\}.$$

We remind the reader that $\omega(e^{i\theta})$ is the unit tangent vector to $\mathbb{D}$ based at the origin and pointing at $e^{i\theta}$, and $\underline{w}^{(2k_0)}$ is defined by (4.1).

We will use the local finiteness of μ to rule this out. The idea is to construct measure preserving maps $\widetilde{\kappa}_i$ s.t.

1. $\widetilde{\kappa}_i$ is well defined on Ω,
2. $\{\widetilde{\kappa}_i(\Omega): i\geq 1\}$ are pairwise disjoint,
3. $\bigcup_{i\geq 1}\widetilde{\kappa}_i(\Omega)\subset\partial\mathbb{D}\times J$ for some compact interval $J\subset\mathbb{R}$.

Since $\widetilde{\kappa}_i$ are measure preserving, $\mu[\partial\mathbb{D}\times J]\geq\mu[\biguplus_{i\geq 1}\widetilde{\kappa}_i(\Omega)]=\infty$, in contradiction to the local finiteness of μ.

Let $s_0:=\max I$ and fix $n_0\in\mathbb{N}$. Given i, follow the geodesic ray of $g^{s_0}[\Gamma\omega(e^{i\theta})]$ until it reaches the in_0–th pop it crosses. "Paste" at the beginning of its cutting sequence there a word of the form $u\underline{w}^{(2k_0)}v$ (where u,v are single letters chosen to ensure reducibility of the concatenation with the existing cutting sequence). This procedure produces a Γ–element $\varphi_{e^{i\theta}}$. Let

$$\widetilde{\kappa}_i(e^{i\theta},s) := (\varphi_{e^{i\theta}}(e^{i\theta}), s-\log|\varphi'_{e^{i\theta}}(e^{i\theta})|). \tag{$*$}$$

- $\widetilde{\kappa}_i$ *is invertible on its image:* To invert, pick some $(e^{i\widetilde{\theta}},\widetilde{s})=\widetilde{\kappa}(e^{i\theta},s)$, and follow the geodesic ray of $\Gamma\omega(e^{i\widetilde{\theta}})$ until the first pop Y where the cutting sequence contains $\underline{w}^{(2k_0)}$. This is the pop where we performed the pasting.[2] The cut'n'paste isometry which cuts the word of the form $u\underline{w}^{(2k_0)}v$ is $\varphi_{e^{i\theta}}^{-1}$. Now that we have identified $\varphi_{e^{i\theta}}$, it is easy to calculate $(e^{i\theta},s)$ from $(e^{i\widetilde{\theta}},\widetilde{s})$.
- $\widetilde{\kappa}_i$ *is measure preserving on its image:* Since $\varphi_{e^{i\theta}}\in\Gamma$, and Γ is countable, $\widetilde{\kappa}_i$ is piecewise measure preserving. Since $\widetilde{\kappa}_i$ is invertible, $\widetilde{\kappa}_i$ is measure preserving.

[2] This is less obvious than it looks, because when we modified the cutting sequence at Y_{in_0}, we may have also inadvertently modified the cutting sequences at Y_j for j close to in_0. We can deal with the problem by taking k_0 and n_0 to be sufficiently large, see [Sa2] for details.

- *The images of $\widetilde{\kappa}_i$ are disjoint:* They can be distinguished from one another by the time the geodesic ray of $\omega(e^{i\theta})$ enters the pop where the cutting sequence contains the word $\underline{w}^{(2k_0)}$. (See the footnote on page 46.)
- $\widetilde{\kappa}_i(\Omega) \subset \partial\mathbb{D} \times J$ *for some J compact independent of i.* This is because the translation on the second coordinate is $-\log|\varphi'_{e^{i\theta}}(e^{i\theta})|$, and (see [Sa2])

$$\begin{aligned}\big|-\log|\varphi'_{e^{i\theta}}(e^{i\theta})|\big| = |R(\varphi_{e^{i\theta}}, e^{i\theta})| \approx |B(\gamma,\gamma')| \le p_Y(\underline{w}^{(2k_0)}) + \text{const.}\\ \le \sup\{p_Y(\underline{w}^{(2k_0)}) : \text{boundary lengths in } [L^{-1}, L]\} + \text{const.} < \infty.\end{aligned}$$

So the statement is proved. □

If (B) holds a.s. then scenario II holds a.s.: Similar idea. We omit the details.

4.5.3 Proof of the holonomy lemmas

We sketch the proof of the second holonomy lemma under the assumption that scenario I holds almost everywhere. A similar argument can be used to prove the first holonomy lemma, and the modifications needed for the scenario II are routine.

Fix $c > 0$. We have to construct an a.e. well defined bijection $\kappa : \partial\mathbb{D} \to \partial\mathbb{D}$ s.t. for a.e. $(e^{i\theta}, s)$, $\kappa(e^{i\theta}) \in \Gamma e^{i\theta}$, $\kappa(e^{i\theta}) \approx e^{i\theta}$, s.t. $-\log|\kappa'(e^{i\theta})|$ is δ–far from $c\mathbb{Z}$ for $\delta = \delta(c)$ and $\big|\log|\kappa'(e^{i\theta})|\big| \le M$ for $M = M(c)$. Fix n_0, T_0, k_0 large.

Fix $e^{i\theta} \in \partial\mathbb{D}$ s.t. $(e^{i\theta}, s)$ satisfies scenario I. Let $\omega(e^{i\theta})$ be the unit vector at $0 \in \mathbb{D}$ which points at $e^{i\theta}$.

1. Follow the geodesic ray γ of $\Gamma\omega(e^{i\theta})$ T_0–units of time, and then some, until the first time it crosses a pop Y where the cutting sequence contains $\underline{w}^{2k}$ with $k > k_0$ and $\underline{w} \in \mathscr{W}(c/n_0, L)$ s.t. $\operatorname{dist}(p_Y(\underline{w}), (c/n_0)\mathbb{Z}) > \delta$ with $\delta = \delta(c/n_0)$ as in Lemma 4.6.
2. Identify the first place in the cutting sequence there where this happens.
3. Let $\varphi_{e^{i\theta}}$ be the cut'n'paste isometry which changes $\underline{w}^{2k}$ to $\underline{w}^{2k+n_0}$.
4. Set $\kappa(e^{i\theta}) := \varphi_{e^{i\theta}}(e^{i\theta})$.

This defines a measurable map $\kappa : \partial\mathbb{D} \cap \{e^{i\theta} : (e^{i\theta}, s) \text{ satisfies scenario I}\} \to \partial\mathbb{D}$. Notice that $\kappa(e^{i\theta})$ is defined for μ-almost everywhere $(e^{i\theta}, s)$.

This map is a bijection. To invert, identify the first place where the cutting sequence contains $\underline{w}^{2k}$ with $k > k_0 + \frac{1}{2}n_0$ and $\underline{w} \in \mathscr{W}(c/n_0, L)$, read $\underline{w}$, "cut" $\underline{w}^{n_0}$. The resulting Γ–element is $\varphi^{-1}_{e^{i\theta}}$. [This argument works if n_0 is large enough, see the footnote on page 46.]

The map κ does not move $e^{i\theta}$ much, because we waited T_0 units of time before making the modification, so the geodesics γ and its $\varphi_{e^{i\theta}}$–modification stay within distance $\sup\{\operatorname{diam}(Y) : Y \in \mathscr{Y}\}$ up to time T_0. So $|\kappa(e^{i\theta}) - e^{i\theta}| = O(e^{-T_0})$.

Let γ denote the projection of the geodesic ray of $\omega(e^{i\theta})$ to $\Gamma \setminus \mathbb{D}$. Let γ' be the $\varphi_{e^{i\theta}}$–modification of γ, then $-\log|\varphi'_{e^{i\theta}}(e^{i\theta})| = R(\varphi_{e^{i\theta}}, e^{i\theta}) \approx B(\gamma,\gamma') \approx n_0 p_Y(\underline{w})$.

Since $p_Y(\underline{w})$ is appproximately δ–far from $\frac{c}{n_0}\mathbb{Z}$, $n_0 p_Y(\underline{w})$ is approximately $n_0\delta$–far from $c\mathbb{Z}$.

A similar argument shows that $\left|\log|\varphi'_{e^{i\theta}}(e^{i\theta})|\right|$ is uniformly bounded: This quantity is approximately bounded by $n_0 p_Y(\underline{w})$ where $\underline{w}$ ranges of the finite set $\mathscr{W}(c/n_0, L)$. The number $p_Y(\underline{w})$ depends continuously on the lengths of the boundary components of Y. Since the triplet of boundary lengths of Y belongs to the compact set $[L^{-1}, L]^3$, $\sup\{p_Y(\underline{w}) : Y \in \mathscr{Y}\} < \infty$. □

4.6 Summary

- Every horocycle e.i.r.m. on $T^1(\Gamma\setminus\mathbb{D})$ lifts to a measure m on $T^1\mathbb{D}$ of the form $d\mu(e^{i\theta}, s)dt$ in *kan*–coordinates, where μ is an e.i.r.m. for the action

$$\varphi(e^{i\theta}, s) = (\varphi(e^{i\theta}), s - \log|\varphi'(e^{i\theta})|) \quad (\varphi \in \Gamma).$$

- $H_\mu := \{s \in \mathbb{R} : \mu \circ g^s \sim \mu\}$ is equal to $\{0\}, c\mathbb{Z}$, or $\mathbb{R}$, and $\exists u : \partial\mathbb{D} \to \mathbb{R}$ Borel s.t. μ is carried by $\{(e^{i\theta}, s) : s \in u(e^{i\theta}) + H_\mu\}$.
- So $\{(e^{i\theta}, s) : s \in u(e^{i\theta}) + H_\mu\}$ is invariant under all measure preserving maps
- Using the assumption that M is very tame, we construct measure preserving maps which do not leave the set $\{(e^{i\theta}, s) : s \in u(e^{i\theta}) + c\mathbb{Z}\}$ invariant. So $H_\mu \neq \{0\}, c\mathbb{Z}$.
- So $H_\mu = \mathbb{R}$, whence $\mu \circ g^s \sim \mu$ for all $s \in \mathbb{R}$. By Babillot's theorem, $dm = e^{\alpha s} d\nu(e^{i\theta}) ds dt$.
- Γ–invariance forces $\frac{d\nu\circ\varphi}{d\nu} = |\varphi'|^\alpha$ which is the same as saying that ν is the boundary values of a Γ–invariant positive eigenfunction

$$\widetilde{F}(z) = \int_{\partial\mathbb{D}} P(e^{i\theta}, z)^\alpha d\nu(e^{i\theta}). \tag{4.2}$$

  So $F(\Gamma z) = \widetilde{F}(z)$ is a positive eigenfunction on $\Gamma\setminus\mathbb{D}$ with eigenvalue $\alpha(\alpha - 1)$. By Sullivan theory and very tameness, $\alpha \geq \frac{1}{2}$ [Sa2].
- Since m is ergodic, ν is extremal in the cone of conformal measures with parameter α. The extremality of ν implies the minimality of F (this uses the uniqueness of the Karpelevich representation when $\alpha \geq \frac{1}{2}$).

So every e.i.r.m. of the horocycle flow arises from a minimal positive eigenfunction of the Laplacian.

4.7 Notes and references

The main reference for this chapter is [Sa2]. Theorem 4.1 was proved by Babillot in [Ba] for Fuchsian groups with Poincaré exponent larger than or equal to $1/2$

(a related result appears in [ASS]). Tame surfaces have such Fuchsian groups by [Sa2]. The idea to prove quasi-invariance by showing that the group H_μ is big is taken from [ANSS]. The cocycle reduction theorem was proved in [Sa1] (see [Rau] for extensions to cocycles taking values in non-abelian groups). §4.1 follows [Sa1] and [LS2]. Busemann's cocycle and Theorem 4.3 were introduced in [Sa2], which is also the reference for the holonomy lemmas for tame surfaces. Special cases of these lemmas were proved before for Fuchsian groups associated to $\mathbb{Z}^d$-covers of compact surfaces [Sa1], and for general regular covers of surfaces of finite area [LS2].

Appendix A
Busemann's function

Busemann's function. Suppose $z, w \in \mathbb{D}$ and $e^{i\theta} \in \partial\mathbb{D}$. *Busemann's function* $b_{e^{i\theta}}(z, w)$ is the signed hyperbolic distance from $\mathrm{Hor}_z(e^{i\theta})$ to $\mathrm{Hor}_w(e^{i\theta})$: the solution s to the equation $g^s[\mathrm{Hor}_z(e^{i\theta})] = \mathrm{Hor}_w(e^{i\theta})$].

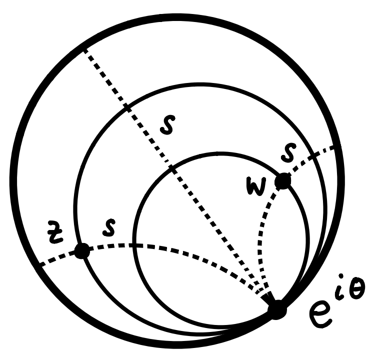

Fig. A.1 $b_{e^{i\theta}}(z, w) = s$

Theorem A.1 (The basic identity for Busemann's function). *For every* $\varphi \in$ *Möb*$(\mathbb{D})$, $b_{\varphi(e^{i\theta})}(0, \varphi(0)) = -\log|\varphi'(e^{i\theta})|$.

Proof. The following properties are obvious:

(I) $b_{\varphi(e^{i\alpha})}(\varphi(z_1), \varphi(z_2)) = b_{e^{i\alpha}}(z_1, z_2)$ for all hyperbolic isometries φ (orientation reversing included)

(II) $b_{e^{i\alpha}}(z_1,z_2)+b_{e^{i\alpha}}(z_2,z_3)=b_{e^{i\alpha}}(z_1,z_3)$
(III) $(e^{i\theta},z,w)\mapsto b_{e^{i\theta}}(z,w)$ is Borel measurable.

We claim that (I), (II), and (III) determine $(e^{i\theta},z,w)\mapsto b_{e^{i\theta}}(z,w)$ up to a multiplicative constant. Suppose $c_{e^{i\theta}}(z,w)$ satisfies (I), (II) and (III).

Firstly, $c_{e^{i\theta}}(z,z)=0$ for all z, because of (II).

Secondly, $c_{e^{i\theta}}(z,w)=0$ whenever $w\in \mathrm{Hor}_{e^{i\theta}}(z)$. To see this, let y denote the midpoint of the horocyclic arc connecting z to w, and let γ denote the geodesic from y to $e^{i\theta}$. Let φ denote the hyperbolic reflection w.r.t. γ,[1] then $\varphi(e^{i\theta})=e^{i\theta}$ and $\varphi(z)=w$, $\varphi(w)=z$. By (I) and (II), $0=c_{e^{i\theta}}(z,w)+c_{e^{i\theta}}(w,z)=c_{e^{i\theta}}(z,w)+c_{\varphi(e^{i\theta})}(\varphi(w),\varphi(z))=2c_{e^{i\theta}}(z,w)$, proving that $c_{e^{i\theta}}(z,w)=0$.

Thirdly, $c_{e^{i\theta}}(z,w)$ is determined by the values of the function $c_1(0,t)$ for t real. To see this use a Möbius transformation to map $e^{i\theta},z$ to $1,0$. Let w^* denote the image of w, and let t denote the intersection of $\mathrm{Hor}_1(w)$ with the real line. Then $c_{e^{i\theta}}(z,w)=c_1(0,w^*)=c_1(0,t)+c_1(t,w^*)=c_1(0,t)$.

Finally, $c_1(0,t)=\mathrm{const.}\,\mathrm{dist}(0,t)$ $(t\in\mathbb{R})$ because (I) implies that $c_1(t_1,t_2)$ is a function of the hyperbolic distance between t_1,t_2, (II) says that this dependence is additive, and (III) says it is Borel.

Here is a construction of a function $c_{e^{i\alpha}}(z,w)$ which satisfies (I), (II) and (III): Let λ_z denote the *harmonic measure* on $\partial\mathbb{D}$ at z, defined by $d\lambda_z(e^{i\theta})=P(e^{i\theta},z)d\theta$, where $P(e^{i\theta},z)=\frac{1-|z|^2}{|e^{i\theta}-z|^2}$ (Poisson's kernel). We claim that

$$c_{e^{i\theta}}(z,w):=\log\frac{d\lambda_z}{d\lambda_w}(e^{i\theta})=\log\left(\frac{P(e^{i\theta},z)}{P(e^{i\theta},w)}\right)$$

satisfies (I),(II), and (III).

(III) is obvious. (II) is the chain rule for Radon-Nikodym derivatives. To see (I) we recall that Poisson's formula says that for every $f\in C(\partial\mathbb{D})$, $F(z):=\int_{\partial\mathbb{D}}fd\lambda_z$ is the unique harmonic function on $\mathbb{D}$ with boundary values $f(e^{i\theta})$. For every $f\in C(\partial\mathbb{D})$ and $\varphi\in\mathrm{Möb}(\mathbb{D})$ $\int fd\lambda_z\circ\varphi^{-1}=\int f\circ\varphi d\lambda_z=F(\varphi(z))=\int_{\partial\mathbb{D}}fd\lambda_{\varphi(z)}$, so $\lambda_z\circ\varphi^{-1}=\lambda_{\varphi(z)}$. This implies (I):

$$c_{\varphi(e^{i\theta})}(\varphi(z),\varphi(w))=\log\frac{d\lambda_z\circ\varphi^{-1}}{d\lambda_w\circ\varphi^{-1}}[\varphi(e^{i\theta})]=\log\left(\frac{d\lambda_z}{d\lambda_w}\circ\varphi^{-1}\right)[\varphi(e^{i\theta})]=c_{e^{i\theta}}(z,w).$$

By the first part of the proof, $b_{e^{i\theta}}(z,w)=\text{const.}\log\frac{P(e^{i\theta},z)}{P(e^{i\theta},w)}$. Since

- $c_1(0,r)=\log\left(\frac{P(1,0)}{P(1,r)}\right)=-\log P(1,r)=-\log\frac{1-r^2}{(1-r)^2}=\log\frac{1-r}{1+r}$,
- $b_1(0,r)=\int_0^r\frac{2dt}{1-t^2}dt=\int_0^r\left(\frac{1}{1-t}+\frac{1}{1+t}\right)dt=\log\frac{1+r}{1-r}$,

the constant equals (-1). We obtain the identity $b_{e^{i\theta}}(z,w)=-\log\left(\frac{d\lambda_z}{d\lambda_w}(e^{i\theta})\right)$.

[1] To construct φ find a Möbius tranformation $\psi:\mathbb{D}\to\mathbb{H}$ which maps $e^{i\theta}$ to ∞ and y to i. Then $\{z,w\}$ map to $\{1-it,1+it\}$. Now reflect using $z\mapsto-\bar{z}$, and go back using ψ^{-1}.

It follows that $b_{\varphi(e^{i\theta})}(0,\varphi(0)) = -\log \frac{d\lambda_0}{d\lambda_{\varphi(0)}}(e^{i\theta}) = \log \frac{d\lambda_0 \circ \varphi^{-1}}{d\lambda_0}[\varphi(e^{i\theta})]$. Since λ_0 is Lebesgue's measure, this equals $\log|(\varphi^{-1})'(\varphi(e^{i\theta}))| = -\log|\varphi'(e^{i\theta})|$. □

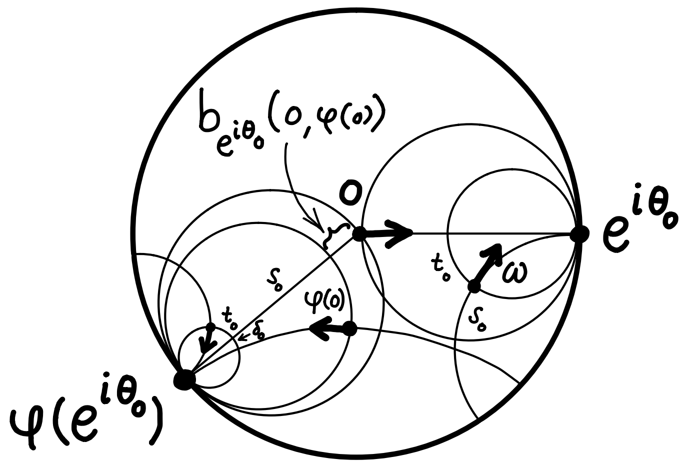

Fig. A.2 Proof of $(*)$

Proof of Theorem 1.1 part (3): Fix $\varphi \in \text{Möb}(\mathbb{D})$, we have to show that

$$\varphi(e^{i\theta_0}, s_0, t_0) = (\varphi(e^{i\theta_0}), s_0 - \log|\varphi'(e^{i\theta_0})|, t_0 + \text{something independent of } t_0) \quad (*)$$

Draw in $\mathbb{D}$ $\omega = (h^{t_0} \circ g^{s_0})[\omega(e^{i\theta_0})]$ together with $\text{Hor}(\omega)$ and $\text{Hor}(\omega(e^{i\theta_0}))$. Add to the picture the geodesic rays of $\omega(e^{i\theta_0})$ and ω. Now draw the image of these figures by φ (Figure A.2).

The *kan*-coordinates of $\varphi(\omega)$ are $(\varphi(e^{i\theta_0}), s_0 + b_{\varphi(e^{i\theta_0})}(0,\varphi(0)), t_0 + \delta_0)$, where δ_0 is some function of $0, \varphi(0), s_0, e^{i\theta_0}$. $(*)$ follows from the basic identity for the Busemann function. □

Proof of Theorem 4.3. The theorem asserts that if $\varphi \in \Gamma$, $e^{i\theta} \in \partial\mathbb{D}$, and $|\varphi(e^{i\theta}) - e^{i\theta}| < 1$, then $|R(\varphi, e^{i\theta}) - B(\gamma, \widetilde{\gamma})| \leq 4|e^{i\theta} - e^{i\theta}|^2$, where

- $R(\varphi, e^{i\theta}) = -\log|\varphi'(e^{i\theta})|$ (the Radon-Nikodym cocycle)
- γ :=the projection to $\Gamma \setminus \mathbb{D}$ of $\gamma[-e^{i\theta}, e^{i\theta}]$, the $\mathbb{D}$-geodesic from $-e^{i\theta}$ to $e^{i\theta}$
- $\widetilde{\gamma}$:=the projection to $\Gamma \setminus \mathbb{D}$ of $\gamma[-e^{i\theta}, \varphi(e^{i\theta})]$, the geodesic from $-e^{i\theta}$ to $\varphi(e^{i\theta})$

◦ $B(\gamma,\widetilde{\gamma}) = \mathrm{dist}_{\widetilde{\gamma}}(\omega_2,\omega_2^*) - \mathrm{dist}_{\gamma}(\omega_1,\omega_1^*)$ for some (any) $\omega_1,\omega_1^* \in \gamma$, $\omega_2,\omega_2^* \in \widetilde{\gamma}$, s.t. $\mathrm{dist}(g^{-s}\omega_1, g^{-s}\omega_2) \xrightarrow[s\to\infty]{} 0$ and $\mathrm{dist}(g^{s}\omega_1^*, g^{s}\omega_2^*) \xrightarrow[s\to\infty]{} 0$.

We take $\omega_1 = \omega_1^*$ =vector based at 0 and pointing at $e^{i\theta}$, ω_2 :=intersection of $\gamma[-e^{i\theta},\varphi(e^{i\theta})]$ and $\mathrm{Hor}_{-e^{i\theta}}(0)$ and ω_2^* :=intersection of $\gamma[-e^{i\theta},\varphi(e^{i\theta})]$ and $\varphi[\mathrm{Hor}_{e^{i\theta}}(0)] = \mathrm{Hor}_{\varphi(e^{i\theta})}(\varphi(0))$ (Figure A.3). Add to the picture ω_3 :=intersection of $\gamma[-e^{i\theta},\varphi(e^{i\theta})]$ and $\mathrm{Hor}_{\varphi(e^{i\theta})}(0)$.

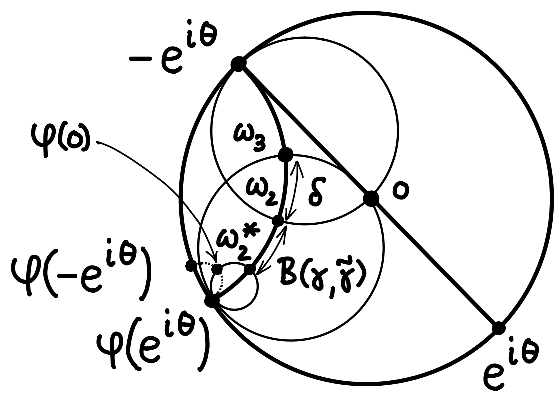

Fig. A.3 Proof of the Approximation Theorem

Clearly $B(\gamma,\widetilde{\gamma}) = \mathrm{dist}_{\gamma[-e^{i\theta},\varphi(e^{i\theta})]}(\omega_2,\omega_2^*) = \mathrm{dist}(\omega_3,\omega_2^*) - \mathrm{dist}(\omega_3,\omega_2)$. The first summand is the signed distance from $\mathrm{Hor}_{\varphi(e^{i\theta})}(0)$ and $\mathrm{Hor}_{\varphi(e^{i\theta})}(\varphi(0))$. This is $b_{\varphi(e^{i\theta})}(0,\varphi(0)) = -\log|\varphi'(e^{i\theta})| = R(\varphi,e^{i\theta})$. So

$$|R(\varphi,e^{i\theta}) - B(\gamma,\widetilde{\gamma})| \le \mathrm{dist}(\omega_3,\omega_2) =: \delta.$$

To estimate δ, let $\vartheta : \mathbb{D} \to \mathbb{H}$ be the Möbius map which sends $e^{i\theta} \mapsto 0$, $-e^{i\theta} \mapsto \infty$, and $0 \mapsto i$. This map maps $\mathrm{Hor}_{-e^{i\theta}}(0)$ to the horizonal line $y = 1$, and $\mathrm{Hor}_{\varphi(e^{i\theta})}(0)$ to a circle passing through i which is tangent to the real axis at $\vartheta(\varphi(e^{i\theta}))$. So δ is the hyperbolic distance between the peak of this circle and $y = 1$ (Fig A.4). It is clear from Figure A.4 that $\delta = O(|\vartheta(e^{i\theta}) - \vartheta[\varphi(e^{i\theta})]|^2)$. A precise calculation using an explicit formula for ϑ shows that $|\delta| \le 4|e^{i\theta} - \varphi(e^{i\theta})|^2$. □

Notes and references. The proof of Theorem A.1 is taken from [Kai]. The proof of Theorem 4.3 is taken from [Sa2].

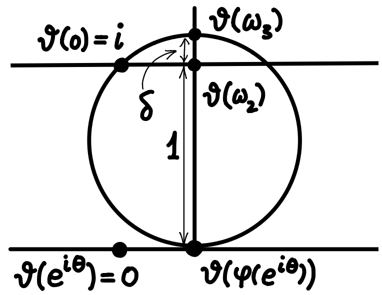

Fig. A.4 Proof of the Approximation Theorem

Appendix A
The cocycle reduction theorem

A.1 Preliminaries on countable equivalence relations

Suppose $(\Omega, \mathscr{F})$ is a standard Borel space (a complete and separable metric space equipped with the σ-algebra of Borel sets). Every measurable group action on Ω generates an equivalence relation

$$x \sim y \Leftrightarrow x, y \text{ are in the same orbit.}$$

This is called the *orbit relation* of the action.

The orbit relation keeps information on the orbits as sets, but forgets the way these sets are parametrized by the group. The language of equivalence relations, which we review below, is designed to handle dynamical properties such as invariance or ergodicity, which only depend on the structure of orbits as unparameterized sets. We will comment on why this is useful at the end of the section.

Countable Borel equivalence relations: These are subsets $\mathfrak{G} \subset \Omega \times \Omega$ such that

1. $x \sim y \Leftrightarrow (x, y) \in \mathfrak{G}$ is a reflexive, symmetric, and transitive relation;
2. the equivalence classes of $\sim$ are all finite or countable;
3. $\mathfrak{G}$ is in the product σ-algebra $\mathscr{F} \otimes \mathscr{F}$.

For example, suppose G is a countable group of bi-measurable maps $g : \Omega \to \Omega$. The the *orbit relation of* G is the countable Borel equivalence relation

$$\mathfrak{orb}(G) := \{(x, g(x)) : x \in X, g \in G\}.$$

Theorem A.1 (Feldman & Moore). *Every countable Borel equivalence relation on a standard measurable space* $(X, \mathscr{F})$ *is the orbit relation of some countable group action on* X.

Corollary A.1. *Suppose* $\mathfrak{G}$ *is a countable Borel equivalence relation on a standard measurable space* $(\Omega, \mathscr{F})$.

1. If $B \in \mathscr{F}$, *then* $\mathrm{Sat}(B) := \{x \in \Omega : \exists y \in B \text{ s.t. } (x, y) \in \mathfrak{G}\}$ *is measuable.*

2. *If* $P \in \mathscr{F} \otimes \mathscr{F}$*, then* $\{x \in \Omega : (x,y) \in \mathfrak{G} \Rightarrow (x,y) \in P\}$ *is measurable.*

Proof. Use the Feldman–Moore Theorem to realize $\mathfrak{G}$ as an orbit relation of a countable group G. Then $\mathrm{Sat}(B) = \bigcup_{g \in G} g(B) \in \mathscr{F}$, and

$$\{x \in \Omega : (x,y) \in \mathfrak{G} \Rightarrow (x,y) \in P\} = \bigcap_{g \in G} \{x \in \Omega : (x, g(x)) \in P\}.$$

This set is measurable because G is countable, and $x \mapsto (x, g(x))$ is measurable. □

"Almost everywhere in $\mathfrak{G}$": Let $P(x,y)$ be a measurable property of pairs $(x,y) \in X \times X$, i.e. $\{(x,y) : P(x,y) \text{ holds}\} \in \mathscr{F} \otimes \mathscr{F}$.

Suppose μ is a measure on X. We say that *P holds μ-a.e. in* $\mathfrak{G}$, if $\{x \in X : (x,y) \in \mathfrak{G} \Rightarrow (x,y) \in P\}$ has full measure. The previous corollary guarantees measurability.

Holonomies, invariant functions, invariant measures: Suppose $\mathfrak{G}$ is a countable Borel equivalence relation.

- A $\mathfrak{G}$–*holonomy* is a bi-measurable bijection $\kappa : A \to B$ where $\mathrm{dom}(\kappa) := A$, $\mathrm{im}(\kappa) := B$ are measurable sets and $(x, \kappa(x)) \in \mathfrak{G}$ for all $x \in \mathrm{dom}(\kappa)$.
- A function $f : \Omega \to \mathbb{R}$ is called $\mathfrak{G}$–*invariant*, if $f \circ \kappa|_{\mathrm{dom}(\kappa)} = f|_{\mathrm{dom}(\kappa)}$ for all $\mathfrak{G}$–holonomies κ. Equivalently, $f(x) = f(y)$ whenever $(x,y) \in \mathfrak{G}$.
- A (possibly infinite) measure μ on Ω is called $\mathfrak{G}$–*invariant* if $\mu \circ \kappa|_{\mathrm{dom}(\kappa)} = \mu|_{\mathrm{dom}(\kappa)}$ for all $\mathfrak{G}$–holonomies κ.
- A $\mathfrak{G}$-invariant measure is called *ergodic*, if every measurable $\mathfrak{G}$–invariant function is equal a.e. to a constant function.

Lemma A.1. *Suppose G is a countable group acting measurably on $(\Omega, \mathscr{F})$, and let μ be a (possibly infinite) measure on $(\Omega, \mathscr{F})$. Let $\mathfrak{G} := \mathfrak{orb}(G)$, then*

1. *μ is G–invariant iff μ is $\mathfrak{G}$–invariant.*
2. *μ is G–ergodic iff μ is $\mathfrak{G}$–ergodic.*

The proof is easy and we leave it to the reader.

Induced equivalence relations: Suppose B is a measurable set with positive measure. The *induced relation on B* is

$$\mathfrak{G}[B] := \mathfrak{G} \cap (B \times B) = \{(x,y) \in B \times B : (x,y) \in \mathfrak{G}\}$$

Lemma A.2. *Suppose μ is a measure on Ω and $\mu(B) > 0$. If μ is $\mathfrak{G}$–invariant, then $\mu|_B$ is $\mathfrak{G}[B]$–invariant. If μ is $\mathfrak{G}$–ergodic, then $\mu|_B$ is $\mathfrak{G}[B]$–ergodic.*

Proof. The first statement is trivial, so we prove the second. Suppose $f : B \to \mathbb{R}$ is $\mathfrak{G}[B]$–invariant. The saturation of B is a $\mathfrak{G}$–invariant measurable set of positive measure (because it contains B). By ergodicity, $\mathrm{Sat}(B)$ has full measure. Define

$$F(x) := \begin{cases} f(y) & \text{for some (any) } y \in B \text{ s.t. } (x,y) \in \mathfrak{G}, \text{ provided } x \in \mathrm{Sat}(B), \\ -666 & \text{whenever } x \notin \mathrm{Sat}(B). \end{cases}$$

The definition is proper because f is $\mathfrak{G}[B]$–invariant. Clearly F is $\mathfrak{G}$–invariant. By $\mathfrak{G}$–ergodicity, F is equal a.e. on Ω to a constant function. So $f = F|_B$ is equal a.e. on B to a constant function. $\square$

Cocycles and skew-product extensions: Suppose $\mathfrak{G}$ is a countable Borel equivalence relation on a standard Borel space $(X, \mathscr{F})$.

- A $\mathfrak{G}$–*cocycle* is a measurable map $\Phi : \mathfrak{G} \to \mathbb{R}$ s.t. $\Phi(x,y) + \Phi(y,z) = \Phi(x,z)$ for all $(x,y),(y,z) \in \mathfrak{G}$. Necessarily $\Phi(x,x) = 0$ and $\Phi(x,y) = -\Phi(y,x)$.
- The *Φ-extension of* $\mathfrak{G}$ is the equivalence relation on $\Omega \times \mathbb{R}$
$$\mathfrak{G}_\Phi := \{\big((x,t),(y,s)\big) \in (\Omega \times \mathbb{R})^2 : (x,y) \in \mathfrak{G}\ ,\ s - t = \Phi(x,y)\}.$$
- Every $\mathfrak{G}$–holonomy $\kappa : A \to B$ generates a $\mathfrak{G}_\Phi$–holonomy $\kappa_\Phi : A \times \mathbb{R} \to B \times \mathbb{R}$ given by $\kappa_\Phi(x,t) = (\kappa(x), t + \Phi(x, \kappa(x)))$.

Example: Radon-Nikodym extensions. Suppose $\Gamma \subset \text{Möb}(\mathbb{D})$ is countable and discrete. Let $\text{Fix}(\Gamma) := \{z \in \partial\mathbb{D} : \exists g \in \Gamma \setminus \{id\} \text{ s.t. } g(z) = z\}$, and set

$$\Omega := \partial\mathbb{D} \setminus \text{Fix}(\Gamma),$$

together with its Borel subsets. This is a standard Borel space.[1]

Let $\mathfrak{G} := \mathfrak{orb}(\Gamma)$. If $(x,y) \in \mathfrak{G}$ then there exists a unique $g \in \Gamma$ such that $y = g(x)$ (otherwise x is a fixed point of a non-trivial element of Γ). Let

$$\Phi(x,y) := -\log|g'(x)| \text{ for the unique } g \in \Gamma \text{ such that } y = g(x).$$

This is a $\mathfrak{G}$–cocycle, because of the chain rule. Then

$$\mathfrak{G}_\Phi = \{\big((x,t),(y,s)\big) \in (\Omega \times \mathbb{R})^2 : \exists g \in \Gamma \text{ s.t. } y = g(x), s = t - \log|g'(x)|\}.$$

Lemma A.3. *Suppose μ is $\mathfrak{G}_\Phi$–ergodic invariant measure on $X \times \mathbb{R}$, then for every $A, B \in \mathscr{F}$ and $K_1, K_2 \subseteq \mathbb{R}$ compact such that $\mu(A \times K_1), \mu(B \times K_2) > 0$ one can find a $\mathfrak{G}$–holonomy κ such that $\mu[\kappa_\Phi(A \times K_1) \cap (B \times K_2)] > 0$.*

Proof. By the Feldman–Moore Theorem, $\mathfrak{G}$ is the orbit relation of a countable group G. Every $g \in G$ determines a $\mathfrak{G}_\Phi$ holonomy with domain $X \times \mathbb{R}$ via $g_\Phi(x,s) = (g(x), s + \Phi(x, g(x)))$. The set $\bigcup_{g \in G} g_\Phi(A)$ is a measurable $\mathfrak{G}_\Phi$–invariant set, whence a set of full measure. So for some $g \in G$, $\mu[g_\Phi(A \times K_1) \cap (B \cap K_2)] > 0$. $\square$

Why do we need all this general nonsense? The Feldman–Moore Theorem says that any countable equivalence relation is the orbit relation of some measurable action of a countable group. The independent minded reader may wonder what is the point of working in this more abstract setup of equivalence relations, when it is not really more general. There are two main reasons:

[1] Since $\text{Fix}(\Gamma)$ is countable, Ω is a G_δ-subset of $\partial\mathbb{D}$. By Alexandrov's Theorem, there is a metric on Ω which turns it into a complete separable metric space, and whose Borel sets are precisely the intersections of Borel subsets of $\partial\mathbb{D}$ with Ω.

1. The language of equivalence relations is convenient in scenarios when it is easier to decide when two points belong to the same orbit, than it is to find the parametrization of the orbit and calculate the actual group element which maps one to the other. This is the case for horocycle flows: There is a simple geometric criterion for deciding when two unit tangent vectors belong to the same horocycle — their geodesic rays are forward asymptotic. But to calculate the horocyclic time it takes to move from one to the other is much more subtle.
2. Induction: It is difficult to construct the induced action of a group on a subset, especially in cases when the individual elements of the group are not conservative (as is the case for hyperbolic or parabolic Möbius transformations). But as we saw above, it is very easy to induce equivalence relations on subsets. Of course, by Feldman–Moore, the induced orbit equivalence relation is the orbit relation of some other countable group action — but constructing that group explicitly is not easy.

We will use the operation of inducing repeatedly in the proof of the cocycle reduction theorem. This is the reason we need to use countable equivalence relations.

A.2 The cocycle reduction theorem

Let $\mathfrak{G}$ be a countable Borel equivalence relation on a standard Borel space $(X,\mathscr{B})$. Let $\Phi : \mathfrak{G} \to \mathbb{R}$ be a measurable cocycle, and suppose μ is a (possible infinite) $\mathfrak{G}_\Phi$–ergodic invariant measure on $X \times \mathbb{R}$. We assume that μ is *locally finite*: $\mu(X \times K) < \infty$ for all compact sets $K \subset \mathbb{R}$.

A *coboundary* is a cocycle of the form $\partial u(x,y) := u(y) - u(x)$. Two cocycles which differ by a coboundary are called *cohomologous*.

The *a.e. range* of a cocycle is the smallest closed subgroup of $\mathbb{R}$ such that $\Phi(x,y) \in H$ μ-a.e. in $\mathfrak{G}_\Phi$.

Sometimes one can reduce the range of a cocycle Φ by subtracting from it a coboundary. For example, if Φ is a $\mathbb{Z}$–valued $\mathfrak{G}$–cocycle, but $u(x)$ is real-valued, then $\Phi + \partial u$ will be an $\mathbb{R}$-valued cocycle. If we subtract ∂u, then we're back to a $\mathbb{Z}$–valued cocycle.

How much can we reduce the range by subtracting a coboundary? The cocycle reduction theorem says that the best we can do is

$$H_\mu := \{s \in \mathbb{R} : \mu \circ g^s \sim \mu\} \overset{!}{=} \{s \in \mathbb{R} : \mu \circ g^s \not\perp \mu\}.$$

Here $g^s : X \times \mathbb{R} \to X \times \mathbb{R}$ is the flow $g^s(x,t) = (x, t+s)$, $\mu \circ g^s \sim \mu$ means that $\mu(g^s E) = 0 \Leftrightarrow \mu(E) = 0$ for all measurable $E \subset X \times \mathbb{R}$, and $\mu \circ g^s \perp \mu$ means that $\mu \circ g^s$ and μ are carried by disjoint sets. Equality $\overset{!}{=}$ is a consequence of ergodicity: Two ergodic invariant measures of the same countable equivalence relation (equiv. countable group action) are either proportional or they are mutually singular.

Theorem A.2 (Cocycle reduction theorem). *If μ is a locally finite $\mathfrak{G}_\Phi$–ergodic and invariant measure on $X\times\mathbb{R}$, then there exists a Borel function $u: X\to\mathbb{R}$ s.t.*

1. The set $\{(x,t): t\in u(x)+H_\mu\}$ has full μ–measure.
2. $\Phi(x,y)+u(x)-u(y)\in H_\mu$ μ–a.e. in $\mathfrak{G}_\Phi$.
3. H_μ is contained in any closed subgroup of $\mathbb{R}$ with property 1 or with property 2.

So H_μ is the minimal μ–a.e. range of the cocycles which are μ-a.e. cohomologous to Φ.

Caution! The reader should note the subtlety in the quantifier in part 2. The measure μ is a measure on $X\times\mathbb{R}$, not X, **and it is not assumed apriori to be a product measure**. Therefore, although the $\mathbb{R}$–coordinates of $(x,t),(y,s)$ are not mentioned explicitly, they do matter — because of their influence on the support of μ. Think of the case when μ is carried by the graph of a function.

The third part of the cocycle reduction theorem is easy:

Lemma A.4. *Suppose μ is a locally finite $\mathfrak{G}_\Phi$–ergodic invariant measure.*

1. *H_μ is a closed subgroup of $\mathbb{R}$, so $H_\mu=\{0\},c\mathbb{Z}$ or $\mathbb{R}$.*
2. *If $u: X\to\mathbb{R}$ is measurable and H is a closed subgroup of $\mathbb{R}$ s.t. $\{(x,t): t\in u(x)+H\}$ has full measure, then $H\supseteq H_\mu$.*
3. *If $u: X\to\mathbb{R}$ is measurable and H is a closed subgroup of $\mathbb{R}$ s.t. $\Phi+\partial u\in H$ μ–a.e. in $\mathfrak{G}_\Phi$, then $H\supseteq H_\mu$.*

Proof. To see the first part, note that there is no loss of generality in assuming that X is a compact metric space, because by Kuratowski's theorem all standard Borel spaces are measurably isomorphic to such spaces. Now proceed as in the proof of Proposition 4.1.

The second part is done by checking that the support of μ is invariant under g^s for all $s\in H_\mu$.

The third part is done by observing that if $\Phi+\partial u\in H$ a.e. in $\mathfrak{G}_\Phi$, then the function $F: X\times\mathbb{R}\to\mathbb{R}/H$, $F(x,t):=t-u(x)+H$ is $\mathfrak{G}_\Phi$ invariant, therefore μ-a.e. constant. So there exists $c\in\mathbb{R}$ s.t. $\{(x,t): t-u(x)\in c+H\}$ has full measure. Arguing as in part 2, we find that $H\supseteq H_\mu$. □

So H_μ is contained in the a.e-range of every cocycle which cohomologous to Φ. It remain to construct the couboundary which reduces the range to H_μ.

We sketch the proof of the cocycle reduction theorem below. For complete details, see [Sa1].

A.3 The proof in case there are no square holes

A *square hole* is a set $B\times[a,b]$ where $B\in\mathscr{B}$, $\mu(B\times[a,b])=0$, and $\mu(B\times\mathbb{R})\neq 0$.

Lemma A.5. *Under the assumptions of the cocycle reduction theorem, if μ has no square holes, then $\mu \circ g^s \sim \mu$ for all $s \in \mathbb{R}$. (Here $g^s(x,\xi) = (x,\xi + s)$.)*

Proof. All standard Borel spaces are isomorphic to compact metric spaces, so there is no loss of generality in assuming that X is a compact metric space equipped with a metric d.

Assume by way of contradiction that $\exists a \in \mathbb{R}$ s.t. $\mu \circ g^a \not\sim \mu$. Since g^s commutes with κ_Φ for every $\mathfrak{G}$–holonomy κ, $\mu \circ g^a$ is $\mathfrak{G}_\Phi$–ergodic and invariant. Two ergodic measures are either equivalent or they are mutually singular (exercise), so $\mu \circ g^a \perp \mu$. Similarly, $\mu \circ g^{-a} \perp \mu$, and $\mu \perp \overline{\mu} := \mu \circ g^a + \mu \circ g^{-a}$.

Choose $f : X \times \mathbb{R} \to [0,\infty)$ with compact support s.t.

$$\int f d\mu = 1 \ , \ \int f d\overline{\mu} < \frac{1}{4} \ , \ \mu[f \neq 0] < \infty.$$

Since f has compact support, f is uniformly continuous. Choose $\delta > 0$ so that

$$\left.\begin{array}{l} d(x,y) < \delta \\ |s-t| < \delta \end{array}\right\} \Longrightarrow |f(x,t) - f(y,s)| < \frac{1}{4\mu[f \neq 0]} \tag{A.1}$$

We will use the assumption that there are no square holes to find a $\mathfrak{G}$–holonomy $\kappa : A \to B$ with the following properties:

(a) for all $x \in A$, $d(x,\kappa(x)) < \delta$;
(b) for all $x \in A$, $\min\{|\Phi(x,\kappa(x)) - a|, |\Phi(x,\kappa(x)) + a|\} < \delta$;
(c) $A \times \mathbb{R}$ has full measure.

Construction of κ: Divide X into a finite pairwise disjoint collection of sets of diameter less than δ. We will construct κ on each cell U separately in such a way that $\kappa(U) \subset U$. Then we will glue the pieces into one holonomy noting that bijectivity is not destroyed because the partition elements are disjoint.

To get (c), we only need to worry about partition sets U such that $\mu(U \times \mathbb{R}) \neq 0$.

Fix some partition set U such that $\mu(U \times \mathbb{R}) \neq 0$. Let $B(t,r) := (t-r, t+r)$. Since there are no square holes,

$$\mu(U \times B(0,\delta/2)) \neq 0 \text{ and } \mu(U \times B(a,\delta/2)) \neq 0.$$

Since μ is $\mathfrak{G}_\Phi$–ergodic, we can use Lemma A.3 to construct a $\mathfrak{G}$–holonomy κ such that $\mu[\kappa_\Phi(U \times B(0,\delta/2)) \cap (U \times B(a,\delta/2))] > 0$.

Let $A_1 := \text{dom}(\kappa) \cap \kappa^{-1}(U)$ and $B_1 := \kappa(A_1)$, then $\mu(A_1 \times \mathbb{R}) > 0$, $\mu(B_1 \times \mathbb{R}) > 0$ and for all $x \in A_1$,

$$x, \kappa(x) \in U \text{ and } \exists |t| < \tfrac{\delta}{2} \text{ s.t. } t + \Phi(x,\kappa(x)) \in B(a, \tfrac{\delta}{2}).$$

So for all $x \in A_1$, $d(x,\kappa(x)) < \delta$ and $|\Phi(x,\kappa(x)) - a| < \delta$.

If $A_1 \times \mathbb{R}$ has full measure in $U \times \mathbb{R}$ we are done and can continue to another partition element. If $B_1 \times \mathbb{R}$ has full measure in $U \times \mathbb{R}$ then we are also done, because we can use κ^{-1}.

If $A_1 \times \mathbb{R}$ and $B_1 \times \mathbb{R}$ are of positive but non-full measure in $U \times \mathbb{R}$, then we let $\kappa_1 := \kappa$ and construct an extension of κ_1 to a bigger domain inside U as follows. Since there are no square holes, $\mu[(U \setminus A_1) \times B(0, \frac{\delta}{2})], \mu[(U \setminus B_1) \times B(a, \frac{\delta}{2})] \neq 0$. By Lemma A.3, there is a $\mathfrak{G}$–holonomy κ' and sets $A_1' \subset U \setminus A_1$, $B_1' \subset U \setminus B_1$ such that $\mu(A_1' \times \mathbb{R}), \mu(B_1' \times \mathbb{R}) \neq 0$ and

$$\mu\left[\kappa'_\Phi(A_1' \times B(0, \tfrac{\delta}{2})) \cap (B_1' \times B(a, \tfrac{\delta}{2}))\right] > 0.$$

As before $d(x, \kappa'(x)) < \delta$ and $|\Phi(x, \kappa'(x)) - a| < \delta$ for $x \in A'$. Since κ, κ' have disjoint domains and disjoint supports, $\kappa_2 := \kappa_1 \cup \kappa'$ is a well-defined holonomy from a subset of U to U.

It is now a standard matter to proceed by the "method of exhaustion" to show that there exists a holonomy κ_∞ with properties (a),(b) and such that one of $\mathrm{dom}(\kappa_\infty) \times \mathbb{R}$, $\mathrm{im}(\kappa_\infty) \times \mathbb{R}$ has full measure in $U \times \mathbb{R}$. See [Sa1] for details. In first case set $\kappa|_U := \kappa_\infty$. In the second case set $\kappa|_U := \kappa_\infty^{-1}$. Now that we are done defining κ a.e. on U, we move to the next partition element. After finitely many steps, we are done.

Using the holonomy κ to prove the lemma: Let $\kappa_\Phi(x,t) := (\kappa(x), t + \Phi(x, \kappa(x)))$.

- κ_Φ preserves μ, because κ_Φ is a $\mathfrak{G}_\Phi$–holonomy;
- $\min\{|f \circ \kappa_\Phi^{-1} - f \circ g^a|, |f \circ \kappa_\Phi^{-1} - f \circ g^{-a}|\} < \frac{1}{4\mu[f \neq 0]}$, because of (A.1).

So $1 = \int f d\mu = \int f d\mu \circ \kappa_\Phi = \int_{\kappa_\Phi[f \neq 0]} f \circ \kappa_\Phi^{-1} d\mu \leq \int_{\kappa_\Phi[f \neq 0]} (f \circ g^a + f \circ g^{-a}) d\mu + \int_{\kappa_\Phi[f \neq 0]} \frac{2}{4\mu[f \neq 0]} d\mu \leq \int f d\overline{\mu} + \frac{1}{2} < \frac{3}{4}$. This contradiction shows that there can be no a s.t. $\mu \circ g^a \not\sim \mu$. ☐

Proof of the cocycle reduction theorem when there are no holes: The lemma shows that if there are no square holes, then $H_\mu = \mathbb{R}$. In this case the cocycle reduction theorem holds with $u \equiv 0$.

A.4 The proof in case there is a square hole

The proof proceeds by determining the support of μ locally, and then globally:

1. Step 1: There is a window $W := A \times [\alpha, \beta]$ with positive μ-measure such that $A \times [\alpha, \beta] = \{(x, u_0(x)) : x \in A\} \mod \mu$ with $u_0 : A \to [\alpha, \beta]$ measurable.
2. Step 2: $A \times \mathbb{R} = \{(x,t) : t \in u_0(x) + H_\mu\} \mod \mu$.
3. Step 3: $A \times \mathbb{R} = \{(x,t) : t \in u(x) + H_\mu\} \mod \mu$ for $u : X \to \mathbb{R}$ measurable such that $u|_A = u_0$.

The main step is step 1; the other two steps follow from ergodicity and invariance. We will make repeated use of the following fact from measure theory:

Lemma A.6. *There exists a probability measure ν on X and Radon measures μ_x on $\mathbb{R}$ such that for every non-negative measurable and μ–integrable function $f : X \times \mathbb{R} \to [0,\infty)$,*

$$\int f d\mu = \int_X \left(\int_{\{x\}\times\mathbb{R}} f(x,t) d\mu_x \right) d\nu(x).$$

For ν–a.e. $x \in X$, for every $\mathfrak{G}$–holonomy κ with $x \in \mathrm{dom}(\kappa)$, $\mu_{\kappa(x)} \circ \kappa_\Phi = \mu_x$.

Sketch of proof (see [A, Ch. 1], [Fu2, Ch. 2], [Schm, Cor. 6.9]). Fix $\varphi : \mathbb{R} \to (0,1)$ such that $\int \varphi(t) d\mu(x,t) = 1$ (such a function exists by local finiteness). Then $\varphi d\mu$ is a probability measure. Since $X \times \mathbb{R}$ is a standard probability space, we have a fibre decomposition of $\varphi\mu$ by general results in measure theory. Multiplying by $1/\varphi$ we obtain a fibre decomposition for μ. Notice that $\nu(E) \equiv \int_{E\times\mathbb{R}} \varphi(t) d\mu(x,t)$.

Any two fibre decompositions of $\varphi\mu$ agree on a set of full measure, because $\int f\varphi d\mu_x$ is a version of the conditional expectation of f on $\mathscr{B}\otimes\{\varnothing,\mathbb{R}\}$, and $L^1(X\times\mathbb{R})$ is separable.

Let G be a countable group of invertible transformations of X such that $\mathfrak{orb}(G) = \mathfrak{G}$ (see the Feldman–Moore Theorem). Comparing the fibre decomposition of μ to that of $\mu \circ \kappa_\Phi$ for $\kappa \in G$, we find that for for a.e. x, $\mu_{\kappa(x)} \circ \kappa_\Phi = \mu_x$ for all $\kappa \in G$. Since G generates $\mathfrak{G}$, this is the case for every $\mathfrak{G}$–holonomy s.t. $\mathrm{dom}(\kappa) \ni x$. ☐

Step 1: *If μ has a square hole, then there is a set with positive measure $W := A \times [\alpha,\beta]$ and a measurable function $u_0 : A \to [\alpha,\beta]$ such that*

(a) for all $((x,\xi),(y,\eta)) \in \mathfrak{G}_\Phi[W] \equiv \mathfrak{G}_\Phi \cap W^2$, $\Phi(x,y) = u_0(y) - u_0(x)$;
(b) $W = \{(x,u_0(x)) : x \in A\}$ mod μ.

Proof. Let $B \times [a,b]$ be a square hole: $\mu(B\times[a,b]) = 0$, $\mu(B\times\mathbb{R}) \neq 0$. Fix some $s \in \mathbb{R}\setminus[a,b]$ and $0 < \varepsilon < \min\{\frac{1}{6}|a-b|, \frac{1}{2}|s-a|\}$ such that

$$\mu(B\times(s-\varepsilon,s+\varepsilon)) \neq 0.$$

Without loss of generality $s < a$, otherwise change coordinates $(x,\xi) \leftrightarrow (x,-\xi)$.

Using Lemma A.6, choose $B_1 \subseteq B$ s.t. $\mu(B_1\times[a,b]) = 0$, $\mu(B_1\times\mathbb{R}) \neq 0$, and so that for all $x \in B_1$

- $\mu_x \sim \mu_{\kappa(x)} \circ \kappa_\Phi$ for every $\mathfrak{G}[B]$–holonomy κ;
- $\mu_x(\{x\}\times[a,b]) = 0$;
- $\mu_x(\{x\}\times(s-\varepsilon,s+\varepsilon)) > 0$.

Next we choose some $t < s$ and some $A \subset B_1$ such that $\mu(A\times\mathbb{R}) \neq 0$ and so that on top of the three bullets above, every $x \in A$ also satisfies

- $\mu_x(\{x\}\times(t-\varepsilon,t+\varepsilon)) = 0$.

Here is how to do this. Let $t := s - (\frac{a+b}{2} - s) \equiv 2s - \frac{a+b}{2}$. We claim that

$$\mu(B_1\times(t-\varepsilon,t+\varepsilon)) = 0 \tag{A.2}$$

Indeed, if this were not the case, then by ergodicity there would exist some $\mathfrak{G}$-holonomy κ and some $B_1' \subset B_1$ such that $\kappa_\Phi(B_1'\times(t-\varepsilon,t+\varepsilon)) \cap (B\times(s-\varepsilon,s+$

ε)) has positive measure. In this case there would also exist some $B_1'' \subset B_1$ with $\mu(B_1'' \times \mathbb{R}) \neq 0$ such that for all $x \in B''$,

$$\kappa(x) \in B\ ,\ |\Phi(x,\kappa(x)) - (s-t)| < 2\varepsilon.$$

So κ_Φ maps $B_1'' \times (s-\varepsilon, s+\varepsilon)$ into $B \times (2s-t-3\varepsilon, 2s-t+3\varepsilon) \subset B \times [a,b]$. But this is impossible, since κ_Φ is measure preserving, $\mu(B \times [a,b]) = 0$, and

$$\mu(B_1'' \times (s-\varepsilon, s+\varepsilon)) = \int_{B_1''} \mu_x(\{x\} \times (s-\varepsilon, s+\varepsilon)) d\nu(x) > 0.$$

($\nu(B_1'') \neq 0$ because $\mu(B_1'' \times \mathbb{R}) \neq 0$). Now that we know (A.2), the existence of A follows from the fibre decomposition of μ.

Define $a' := t-\varepsilon$, $b' := t+\varepsilon$, and choose $[\alpha,\beta] \subset (s-\varepsilon, s+\varepsilon)$ such that $\mu(A \times [\alpha,\beta]) > 0$ and $|\alpha-\beta| < \frac{1}{3}\varepsilon$. Necessarily

$$|\alpha - \beta| < \frac{1}{2}\min\{|a'-b'|, |a-b|, |a-\beta|, |b'-\alpha|\}.$$

Indeed, $|a'-b'| = 2\varepsilon$, $|a-b| > 6\varepsilon$, $|a-\beta| > a-s-\varepsilon \geq \varepsilon$, and $|b'-\alpha| > (s-\varepsilon)-(t+\varepsilon) = (s-t)-2\varepsilon = (\frac{a+b}{2}-s)-2\varepsilon > (\frac{a+b}{2}-a)-2\varepsilon = \frac{|a-b|}{2}-2\varepsilon \geq \varepsilon$.

We show that $W := A \times [\alpha,\beta]$ satisfies the requirements of step 1. Define

$$U(x) := \inf\{\tau \in [b',b] : \mu_x(\{x\} \times (\tau,b]) = 0\}$$

Recall that $\mathfrak{G}[A] := \mathfrak{G} \cap A^2$, and fix some $\mathfrak{G}[A]$–holonomy κ. If $x \in A \cap \mathrm{dom}(\kappa)$, $x' := \kappa(x) \in A$ and $|\Phi(x,x')| < |\alpha-\beta|$, then

$$\begin{aligned}
U(x') &= \inf\{\tau \in [b',b] : \mu_{\kappa(x)}(\{\kappa(x)\} \times (\tau,b]) = 0\} \\
&= \inf\{\tau \in [\alpha,a] : \mu_{\kappa(x)}(\{\kappa(x)\} \times (\tau, \tfrac{a+b}{2}]) = 0\} \quad \because \begin{pmatrix} \mu_{x'}(\{x'\} \times [\frac{a+b}{2},b]) = 0 \\ \mu_{x'}(\{x'\} \times [\alpha,\beta]) > 0 \end{pmatrix} \\
&= \inf\{\tau \in [\alpha,a] : (\mu_{\kappa(x)} \circ \kappa_\Phi)(\{x\} \times (\tau - \Phi(x,x'), \tfrac{a+b}{2} - \Phi(x,x')]) = 0\} \\
&= \inf\{\tau \in [\alpha,a] : (\mu_{\kappa(x)} \circ \kappa_\Phi)(\{x\} \times (\tau - \Phi(x,x'), a]) = 0\} \quad \because \tfrac{a+b}{2} - |\alpha-\beta| > a \\
&= \inf\{\tau \in [\alpha,a] : \mu_x(\{x\} \times (\tau - \Phi(x,x'), a]) = 0\} \\
&= \inf\{\tau' \in [\alpha - \Phi(x,x'), a - \Phi(x,x')] : \mu_x(\{x\} \times (\tau', a]) = 0\} + \Phi(x,x') \\
&\geq \inf\{\tau' \in [b',b] : \mu_x(\{x\} \times (\tau', a]) = 0\} + \Phi(x,x') = U(x) + \Phi(x,x').
\end{aligned}$$

So $\Phi(x,x') \leq U(x') - U(x)$. Exchanging the places of x, x' and noting that $\Phi(x',x) = -\Phi(x,x')$, we find that $\Phi(x,x') = U(x') - U(x)$.

It follows that the function $F(x,\xi) := \xi - U(x)$ is invariant with respect to the induced equivalence relation $\mathfrak{G}_\Phi[W]$. By Lemma A.2, this equivalence relation is ergodic. So $\xi - U(x) = \mathit{const}$ μ–a.e. in W.

The step follows with $u_0(x) := U(x) + \mathrm{const}$.

Step 2: *Either* $A \times \mathbb{R} = \{(x, u_0(x)) : x \in A\} \mod \mu$ *and* $H_\mu = \{0\}$*, or there exists* $c > 0$ *s.t.* $A \times \mathbb{R} = \{(x, u_0(x) + cn) : x \in A, n \in \mathbb{Z}\} \mod \mu$ *and* $H_\mu = c\mathbb{Z}$*. In both cases,* $\Phi(x,y) + u_0(x) - u_0(y) \in H_\mu$ μ*–a.e. in* $\mathfrak{G}_\Phi[A \times \mathbb{R}]$.

Sketch of proof (see [Sa1] for details). Let

$$u_1(x) := \sup\{t \geq u_0(x) : \mu_x(\{x\} \times (u_0(x), t)) = 0\}.$$

Notice that $\mu_x[\{x\} \times (u_0(x), u_1(x))] = 0$, and if $u_1(x) < \infty$ then

$$\mu_x\left[\{x\} \times [u_1(x), u_1(x) + \varepsilon)\right] > 0 \text{ for all } \varepsilon > 0.$$

Suppose $((x,\xi),(x',\eta)) \in \mathfrak{G}_\Phi[W]$ and let κ be a $\mathfrak{G}$–holonomy such that $(x',\eta) = \kappa_\Phi(x,\xi)$. Since $\mu_{\kappa(x)} \circ \kappa_\Phi = \mu_x$, $u_1(x') < \infty$ iff $u_1(x) < \infty$, and in this case the identity $u_0(x') = u_0(x) + \Phi(x,x')$ implies that

$$\mu_{x'}\left[\{x'\} \times (u_0(x'), u_1(x) + \Phi(x,x'))\right] = 0$$
$$\mu_{x'}\left[\{x'\} \times [u_1(x) + \Phi(x,x'), u_1(x) + \Phi(x,x') + \varepsilon)\right] > 0 \text{ for all } \varepsilon > 0.$$

It follows that $u_1(x') = u_1(x) + \Phi(x,x')$.

Recall that $u_0(x') = u_0(x) + \Phi(x,x')$, then $u_1(x') - u_0(x') = u_1(x) - u_0(x)$, proving that $u_1 - u_0$ is $\mathfrak{G}_\Phi[W]$–invariant. By ergodicity, either $u_1 < \infty$ μ–a.e. in W and then $u_1 = u_0 + \text{const.}$, or $u_1 = \infty$ μ–a.e. in W. Because $\mu|_W \sim \int_A \delta_{(x,u_0(x))} d\nu(x)$, instead of saying "$\mu$–a.e. in W" we can say "ν–a.e. in A." In summary: $u_1 = u_0 + c$ ν–a.e. in A, where $c \in [0,\infty]$.

We claim that $c > 0$. By step 1, for every $x \in A$, μ_x has a single atom in $\{x\} \times [\alpha,\beta]$ (at $(x, u_0(x))$). So $u_0(x) \leq \beta \leq u_1(x)$, and the only way for c to be equal to zero is to have $u_0(x) = u_1(x) = \beta$. If this is the case, then

$$\mu(A \times \{\beta\}) > 0 \text{ and } \mu(A \times (\beta, \beta + \delta)) > 0 \text{ for all } \delta > 0.$$

But then by ergodicity we can find $A' \subset A$ such that $\mu(A \times (\beta, \beta + |\alpha - \beta|)) > 0$ and a $\mathfrak{G}$–holonomy κ such that $\kappa(A') \subset A$ and $|\alpha - \beta| < \Phi(x, \kappa(x)) < 0$ on A'. But this is absurd because in this case κ_Φ maps $A \times \{\beta\}$ into $A \times [\alpha, \beta)$ which has zero measure by the assumption that $u_0 = \beta$ on W.

We now separate cases.

CASE 1: $u_1 < \infty$ ν*–a.e. in* A.

In this case, a similar argument to the one we just used shows that for ν–a.e. every $x \in A$, $\mu_x(\{x\} \times (u_1(x), u_1(x) + \delta)) = 0$ for all δ small enough. So $(x, u_1(x)) = (x, u_0(x) + c)$ is an atom of μ_x. Since $\mu_x(x, u_0(x)) > 0$ and $\mu_x(x, u_0(x) + c) > 0$ for ν–a.e. $x \in A$, μ and $\mu \circ g^c$ are *not* mutually singular. So $c \in H_\mu$.

Similarly, since $\mu_x(\{x\} \times (u_0(x), u_0(x) + c)) = \mu_x(\{x\} \times (u_0(x), u_1(x)) = 0$ for ν–a.e. $x \in A$, $\mu \not\sim \mu \circ g^\tau$ for $0 < \tau < c$. So $H_\mu = c\mathbb{Z}$.

Since $\mu_x(\{x\}\times(u_0(x),u_0(x)+c))=0$, $\mu_x(x,u_0(x))>0$, and $\mu_x(x,u_0(x)+c)>0$ for ν–a.e. $x\in A$, $\mu_x\sim\sum_{k\in\mathbb{Z}}\delta_{(x,u_0(x)+kc)}$ for a.e. $x\in A$. It follows that

$$A\times\mathbb{R}=\{(x,u_0(x)+kc):x\in A,k\in\mathbb{Z}\}\mod\mu.$$

The $\mathfrak{G}_\Phi$–invariance of μ now implies that for a.e. $(x,\xi)\in A\times\mathbb{R}$, for all (countably many) $(y,\eta)\in A\times\mathbb{R}$ s.t. $((x,\xi),(y,\eta))\in\mathfrak{G}_\Phi[A\times\mathbb{R}]$,

$$\xi\in u_0(x)+c\mathbb{Z}\ ,\ \eta\in u_0(y)+c\mathbb{Z}\ ,\ \eta=\xi+\Phi(x,y),$$

whence $\Phi(x,y)+u_0(x)-u_0(y)\in c\mathbb{Z}$. In other words, $\Phi(x,y)+u_0(x)-u_0(y)\in H_\mu$ a.e. in $\mathfrak{G}_\Phi[A\times\mathbb{R}]$.

CASE 2: $u_1=\infty$ *ν–a.e. in A*

In this case $\mu_x(\{x\}\times(u_0(x),\infty))=0$ but $\mu_x(x,u_0(x))>0$ ν-a.e. in A.

We claim that $\mu_x((-\infty,u_0(x))=0$ ν–a.e. in A. The argument is similar to the one we used before, so we only sketch it: Had there been some mass below the graph of u_0 on A, then by ergodicity there would be some $\mathfrak{G}_\Phi$–holonomy which maps a positive measure part of $A\times\mathbb{R}$ into $A\times\mathbb{R}$ in such a way that Φ takes strictly positive values. This holonomy would shift some positive measure piece of the graph of u_0 strictly up in a measure preserving way. But this is impossible because there is no mass above the graph of u_0.

Thus $A\times\mathbb{R}=\{(x,u_0(x)):x\in A\}\mod\mu$. It automatically follows that $H_\mu=\{0\}$. Again, this implies that $\Phi(x,y)+u_0(x)-u_0(y)=0$ a.e. in $\mathfrak{G}_\Phi[A\times\mathbb{R}]$. ☐

Step 3: *There exists $u:X\to\mathbb{R}$ measurable s.t. $X\times\mathbb{R}=\{(x,\xi):\xi\in u(x)+H_\mu\}$ mod μ and $\Phi(x,y)+u(x)-u(y)\in H_\mu$ μ–almost everywhere in $\mathfrak{G}$.*

Proof. Define $F_0:A\to\mathbb{R}/H_\mu$ by $F_0(x,\xi):=u_0(x)+H_\mu$.

We can extend F_0 to $\mathrm{Sat}(A)=\{y\in X:\exists x\in A \text{ s.t. } (x,y)\in\mathfrak{G}\}$ by setting

$$F(y):=F_0(x)+\Phi(x,y)\text{ for some (any) }x\in A\text{ s.t. }(x,y)\in\mathfrak{G}.$$

The definition is proper, because if $x_1,x_2\in A$ both satisfy $(x_i,y)\in\mathfrak{G}$, then

$$\begin{aligned}&[F_0(x_1)+\Phi(x_1,y)]-[F_0(x_2)+\Phi(x_2,y)]\\&=u_0(x_1)-u_0(x_2)+\Phi(x_1,y)+\Phi(y,x_2)+H_\mu\\&=u_0(x_1)-u_0(x_2)+\Phi(x_1,x_2)+H_\mu=H_\mu,\text{ by step 2.}\end{aligned}$$

By construction, $F=F_0$ on A and for every $x\in\mathrm{Sat}(A)$, for every y s.t. $(x,y)\in\mathfrak{G}$,

$$\Phi(x,y)+F(x)-F(y)=H_\mu.$$

Let $C:\mathbb{R}/H_\mu\to\mathbb{R}$ be a measurable (even piecewise continuous) function such that $C(\tau+H_\mu)\in\tau+H_\mu$, and let

$$u(x):=C(F(x)).$$

Then $\Phi(x,y) + u(x) - u(y) \in H_\mu$ a.e. in $\mathfrak{G}_\Phi$.

It immediately follows that $G(x,\xi) := \xi - u(x) + H_\mu$ is $\mathfrak{G}_\Phi$–invariant, whence a.e. constant. The constant is zero because $G = H_\mu$ on the positive measure set $A \times [\alpha, \beta]$. So $\xi - u(x) \in H_\mu$ μ-a.e., whence $X \times \mathbb{R} = \{(x,\xi) : \xi \in u(x) + H_\mu\}$. □

A.5 Notes and references

The cocycle reduction theorem is taken from [Sa1], as is the proof sketched above. Extensions to cocycles taking values in non-abelian groups are given in [Rau].

References

[A] J. Aaronson: An introduction to infinite ergodic theory. *Mathematical Surveys and Monographs*, **50**. *American Mathematical Society, Providence, RI, 1997. xii+284 pp. ISBN: 0-8218-0494-4*

[ADF] J. Aaronson; M. Denker; A.M. Fisher: *Second order ergodic theorems for ergodic transformations of infinite measure spaces.* Proc. Amer. Math. Soc. **114** (1992) no. 1, 115–127.

[ANSS] J. Aaronson; H. Nakada; O. Sarig; R. Solomyak: *Invariant measures and asymptotics for some skew products.* Israel J. Math. **128** (2002), 93–134.

[ASS] Aaronson, J.; Sarig, O.; Solomyak, R.: *Tail-invariant measures for some suspension semiflows.* Discrete Contin. Dyn. Syst. **8** (2002), no. 3, 725735.

[AW] J. Aaronson and B. Weiss: *On the asymptotics of a 1-parameter family of infinite measure preserving transformations.* Bol. Soc. Brasil. Mat. (N.S.) **29** (1998), no. 1, 181–193.

[Ba] M. Babillot: *On the classification of invariant measures for horospherical foliations on nilpotent covers of negatively curved manifolds.* In: *Random walks and geometry* (V.A. Kaimanovich, Ed.) de Gruyter, Berlin (2004), 319–335.

[BL1] M. Babillot, F. Ledrappier: : *Lalley's theorem on periodic orbits of hyperbolic flows.* Ergodic Theory Dynam. Systems **18** (1998), no. 1, 17–39.

[BL2] M. Babillot, F. Ledrappier: *Geodesic paths and horocycle flows on Abelian covers.* Lie groups and ergodic theory (Mumbai, 1996), 1–32, Tata Inst. Fund. Res. Stud. Math. **14**, Tata Inst. Fund. Res., Bombay, (1998).

[Bek] M. Bachir Bekka: *Ergodic theory and topological dynamics of group actions on homogeneous spaces*, London Math. Soc. Lecture Notes Series **269**, Cambridge University Press, 2013.

[BE] Bougerol, Philippe; Élie, Laure: *Existence of positive harmonic functions on groups and on covering manifolds.* Ann. Inst. H. Poincaré Probab. Statist. **31** (1995), no. 1, 59–80.

[Bu] M. Burger: *Horocycle flow on geometrically finite surfaces.* Duke Math. J. **61** (1990), no. 3, 779–803.

[CD] Choquet, Gustave; Deny, Jacques: *Sur l'équation de convolution* $\mu = \mu * \sigma$. (French) C. R. Acad. Sci. Paris **250** 1960 799–801.

[CG] Conze, J.-P.; Guivarc'h, Y.: *Propriété de droite fixe et fonctions harmoniques positives.* (French) Théorie du potentiel et analyse harmonique (Exposés des Journées de la Soc. Math. France, Inst. Recherche Math. Avancée, Strasbourg, 1973), pp. 126–132. Lecture Notes in Math., Vol. **404**, Springer, Berlin, 1974.

[Cou] Y. Coudene: *Cocycles and stable foliations of Axiom A flows*, Ergodic Th. & Dynam. Syst. **21** (2001), 767–774.

[Dal] F. Dal'bo: *Remarques sur le spectre des longueurs d'une surface et comptages.* Bol. Soc. Brasil. Mat. (N.S.) **30** (1999), no. 2, 1991.

[Da] S. G. Dani: *Invariant measures of horospherical flows on noncompact homogeneous spaces.* Invent. Math. **47** (1978), no. 2, 101–138.

[DS] S. G. Dani, J. Smillie: *Uniform distribution of horocycle orbits for Fuchsian groups.* Duke Math. J. **51** (1984), 185–194.

[FM] J. Feldman; C.C. Moore: *Ergodic equivalence relations, cohomology, and von Neumann algebras. I.* Trans. Amer. Math. Soc. **234** (1977), no. 2, 289–324.

[Fi1] A. Fisher: *Convex-invariant means and a pathwise central limit theorem.* Adv. in Math. **63** (1987), no. 3, 213–246.

[Fi2] A.M. Fisher: *Integer Cantor sets and an order-two ergodic theorem.* Ergodic Theory Dynam. Systems **13** (1993), no. 1, 45–64.

[Fu1] H. Furstenberg: *The unique ergodicity of the horocycle flow.* Springer Lecture Notes **318** (1972), 95–15.

[Fu2] Furstenberg, H. Recurrence in ergodic theory and combinatorial number theory. *M. B. Porter Lectures. Princeton University Press, Princeton, N.J., 1981. xi+203 pp.*

[Gui] Guivarc'h, Y.: *Sur la représentation intégrale des fonctions harmoniques et des fonctions propres positives dans un espace riemannien symétrique*. Bull. Sci. Math. (2) **108** (1984), no. 4, 373–392.

[GR] Y. Guivarc'h and A. Raugi: *Products of random matrices: convergence theorems.* In *Random matrices and their applications* (Brunswick, Maine, 1984), 31–54, Contemp. Math., **50**, Amer. Math. Soc., Providence, RI, (1986).

[Ho1] E. Hopf: Ergodentheorie, *Ergeb. Mat. vol.* **5**, *Springer, Berlin, 1937.*

[Ho2] E. Hopf: *Ergodic theory and the geodesic flow on surfaces of constant negative curvature*, Bull. AMS **77** (1971), 863–877.

[Hub] J. H. Hubbard: Teichmüller Theory and applications to geometry, topology, and dynamics. Volume 1: Teichmüller theory. *xx+459 pages. Matrix Edition (2006).*

[Kai] V. A. Kaimanovich: *Ergodic properties of the horocycle flow and classification of Fuchsian groups.* J. Dynam. Control Systems **6** (2000), no. 1, 21–56.

[Kar] F.I. Karpelevich: *The geometry of geodesics and the eigenfunctions of the Laplacian on symmetric spaces.* Trans. Moskov. Math. Soc. **14** 48–185 (1965).

[K] S. Katok: Fuchsian groups. *x+175 pages. The U. of Chicago Press (1992).*

[KS] A. Katsuda and T. Sunada: *Closed orbits in homology classes.* Inst. Hautes Études Sci. Publ. Math. No. **71** (1990), 5–32.

[La] S. Lalley: *Closed geodesics in homology classes on surfaces of variable negative curvature,* Duke Math. J., **55** (1989), 795–821.

[L] F. Ledrappier: *Invariant measures for the stable foliation on negatively curved periodic manifolds*, Ann. Inst. Fourier **58** number 1 (2008), 85–105.

[LS1] F. Ledrappier and O. Sarig: *Unique ergodicity for non-uniquely ergodic horocycle flows.* Discrete Contin. Dyn. Syst. **16** (2006), no. 2, 411–433.

[LS2] F. Ledrappier; O. Sarig: *Invariant measures for the horocycle flow on periodic hyperbolic surfaces.* Israel J. Math. **160**, 281-317 (2007).

[LS3] F. Ledrappier and O. Sarig: *Fluctuations of ergodic sums for horocycle flows on* $\mathbb{Z}^d$*-covers of finite volume surfaces.* Discrete Contin. Dyn. Syst. **22** (2008), no. 1-2, 247–325.

[LP] Lin, V. Ya.; Pinchover, Y.: *Manifolds with group actions and elliptic operators.* Mem. Amer. Math. Soc. **112** (1994), no. 540, vi+78 pp.

[LyS] Lyons, Terry; Sullivan, Dennis: *Function theory, random paths and covering spaces.* J. Differential Geom. **19** (1984), no. 2, 299–323.

[M] G. Margulis: *Positive harmonic functions on nilpotent groups.* Dokl. Akad. Nauk SSSR **166** 1054–1057 (Russian); translated as Soviet Math. Dokl. **7** 1966 241–244.

[Pat] Patterson, S.J.: *The limit set of Fuchsian group*, Acta Math. **136** (1976), 241–273.

[Po] Pollicott, M.: $\mathbb{Z}^d$*-covers of horospheric foliations*, Discrete Continuous Dynam. Syst. **6** (2000), 599–604.

[Ra1] Ratner, M.: *A central limit theorem for $\mathcal{Y}$-flows on three-dimensional manifolds.* (Russian) Dokl. Akad. Nauk SSSR **186** (1969) 519–521.

[Ra2] Ratner, M.: *On Raghunathan's measure conjecture.* Ann. of Math. (2) **134** (1991), no. 3, 545–607.

[Ra3] Ratner, M.: *Raghunathan's topological conjecture and distributions of unipotent flows.* Duke Math. J. **63** (1991), no. 1, 235–280.

[Rau] Raugi, A.: *Mesures invariantes ergodiques pour des produits gauches.* Bull. Soc. Math. France **135** (2007), no. 2, 247–258.

[Ree] Rees, M.: *Divergence type of some subgroups of finitely generated Fuchsian groups.* Ergodic Theory Dynamical Systems **1** (1981), no. 2, 209–221.

[Ro1] Roblin, T,: *Sur l'ergodicité rationnelle et les propriétés ergodiques du flot géodésique dans les variétés hyperboliques.* Ergodic Theory Dynam. Systems **20** (2000), no. 6, 1785–1819.

[Ro2] T. Roblin: *Ergodicité et équidistribution en courbure négative.* Mém. Soc. Math. Fr. (N.S.) **95** (2003), vi+96 pp.

[Sa1] O. Sarig: *Invariant measures for the horocycle flow on Abelian covers.* Inv. Math. **157**, 519–551 (2004).

[Sa2] O. Sarig: *The horocyclic flow and the Laplacian on hyperbolic surfaces of infinite genus.* Geom. Funct. Anal. **19** (2010), no. 6, 1757–1812.

[SSc] O. Sarig; B. Schapira: *The generic points for the horocycle flow on a class of hyperbolic surfaces with infinite genus.* Int. Math. Res. Not. IMRN 2008, Art. ID rnn 086, 37 pp.

[Scha] B. Schapira: *Equidistribution of the horocycles of a geometrically finite surface.* Int. Math. Res. Not. **40**, 2447–2471 (2005).

[Schm] K. Schmidt: Cocycles on ergodic transformation groups. *Macmillan Lectures in Mathematics, Vol. 1. Macmillan Company of India, Ltd., Delhi, 1977. 202 pp*. (Available from the author's homepage.)

[Se1] C. Series: *Geometrical methods of symbolic coding.* In *Ergodic Theory, Symbolic Dynamics, and Hyperbolic Spaces* Edited by T. Bedford, M. Keane, C. Series. Oxford Univ. Press (1991).

[Sol] R. Solomyak: *A short proof of the ergodicity of the Babillot-Ledrappier measures*, Proc. AMS **129** (2001), 3589–3591.

[Sta] A.N. Starkov: *Fuchsian groups from the dynamical viewpoint*, J. Dynamics and Control Systems **1** (1995), 427–445.

[Sti] J. Stillwell: Geometry of surfaces. *Universitext. Springer-Verlag, New York, 1992. xii+216 pp. ISBN: 0-387-97743-0*

[Su1] D. Sullivan: *The density at infinity of a discrete group of hyperbolic motions.* Inst. Hautes Études Sci. Publ. Math. No. **50** (1979), 171–202.

[Su2] D. Sullivan: *Related aspects of positivity in Riemannian geometry.* J. Diff. Geom. **25** 327–351 (1987).

[Z] R. Zweimüller: *Hopf's ratio ergodic theorem by inducing.* Colloq. Math. **101** (2004), no. 2, 289–292.